From the DEPARTMENT OF MOLECULAR MEDICINE & SURGERY
Karolinska Institutet, Stockholm, Sweden

EXPANDING THE GENETIC AND PHENOTYPIC SPECTRUM OF SKELETAL DYSPLASIAS

Anna Hammarsjö

Karolinska Institutet

Stockholm 2018

Published by Karolinska Institutet.
Printed by E-print AB 2018

ISBN 978-91-7831-121-7

EXPANDING THE GENETIC AND PHENOTYPIC SPECTRUM OF SKELETAL DYSPLASIAS

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Anna Hammarsjö

Principal Supervisor:
Associate Professor Giedre Grigelioniene
Karolinska Institutet
Department of Molecular Medicine and Surgery

Co-supervisor(s):
Professor Magnus Nordenskjöld
Karolinska Institutet
Department of Molecular Medicine and Surgery

Professor Ann Nordgren
Karolinska Institutet
Department of Molecular Medicine and Surgery

Associate Professor Anna Lindstrand
Karolinska Institutet
Department of Molecular Medicine and Surgery

Daniel Nilsson, PhD
Karolinska Institutet
Department of Molecular Medicine and Surgery

Opponent:
Professor Geert Mortier
University of Antwerp
Department of Medical Genetics

Examination Board:
Professor Ola Nilsson
Karolinska Institutet
Department of Women's and Children's health/
Örebro University
Department of Medical Sciences
Division of Endocrinology

Associate Professor Ann-Charlotte Thuresson
Uppsala University
Department of Immunology, Genetics and Pathology
Division of Medical Genetics and Genomics

Associate Professor Larry Mansouri
Karolinska Institutet
Department of Molecular Medicine and Surgery
Division of Clinical Genetics

”The rabbit runs faster than the fox,
because the rabbit is running for his life
while the fox is only running for his dinner”

Richard Dawkins, The Selfish Gene

ABSTRACT

Skeletal dysplasias constitute a large and heterogeneous group of disorders, many causing disabilities with profound effects on the quality of life of the affected individuals and their families. Each individual skeletal dysplasia is rare, however, as more than 450 different disorders have been described, skeletal dysplasias as a group affect approximately three in 10 000 individuals. The age of onset for skeletal dysplasias ranges from prenatal to adult, but most of the affected individuals are diagnosed in childhood. Skeletal dysplasias mainly affect bone and cartilage, but symptoms may involve other organs, such as sensorineural hearing loss in Stickler syndrome, nephronophthisis in Sensenbrenner syndrome, and structural heart abnormalities in acromicric dysplasia. Therefore, most skeletal dysplasias can be defined as syndromes with a significant skeletal involvement.

Clinical diagnosis of skeletal dysplasias is based on meticulous phenotypic characterization, skeletal radiography (for pattern recognition) and genetic testing. Molecular diagnostics has improved significantly by massively parallel sequencing (MPS) technologies, such as whole exome and genome sequencing. However, even after extensive clinical phenotyping and advanced molecular analyses, many patients with congenital skeletal disorders still lack molecular diagnoses and many clinical entities are not well-characterized regarding their natural course and complications. Molecular diagnosis is important since it gives information about prognosis and recurrence risk, as well as, in some cases possibilities to offer targeted treatment, participation in clinical trials, and tailored medical follow-up. This thesis focuses on gene discovery, studies of previously clinically defined skeletal dysplasias with unknown genetic background and aims to improve the molecular analyses for patients with diagnoses which are difficult to solve.

In **study I**, we identify a novel pathogenic variant in *ALG9*, as the cause of a lethal skeletal syndrome in two families. In **study II**, we show that a variant in *COL2A1* causes spondyloepiphyseal dysplasia type Stanescu. **Study III** shows that pathogenic variants in *BMPER* cause ischiospinal dysostosis (ISD), which is allelic to diaphanospondylodysostosis. In **study IV**, we describe a novel skeletal ciliopathy in four individuals with spondylometaphyseal dysplasia and thorax hypoplasia caused by pathogenic variants in *KIAA0753*. Finally, **study V** represents a cohort of 24 unrelated patients with skeletal ciliopathies, where we solve the genetic diagnoses in 83% of them. Here, we show two rare intronic variants and two exonic synonymous variants leading to aberrant splicing, which indicates that extended RNA studies are necessary to improve molecular diagnostics in some cases.

Altogether, the results of these studies expand the genetic and phenotypic spectrum of skeletal dysplasias and demonstrate that MPS technology in combination with meticulous phenotyping is a powerful method to discover disease-causing variants in patients with congenital skeletal disorders.

LIST OF SCIENTIFIC PAPERS INCLUDED IN THE THESIS

I. **A novel phenotype in N-glycosylation disorders: Gillessen-Kaesbach-Nishimura skeletal dysplasia due to pathogenic variants in *ALG9*.**
Tham E[*], Eklund EA[*], Hammarsjö A[*], Bengtson P, Geiberger S, Lagerstedt-Robinson K, Malmgren H, Nilsson D, Grigelionis G, Conner P, Lindgren P, Lindstrand A, Wedell A, Albåge M, Zielinska K, Nordgren A, Papadogiannakis N, Nishimura G, Grigelioniene G.
Eur J Hum Genet. 2016 Feb;24(2):198-207.

II. **Pathogenic variant in the *COL2A1* gene is associated with Spondyloepiphyseal dysplasia type Stanescu.**
Hammarsjö A, Nordgren A, Lagerstedt-Robinson K, Malmgren H, Nilsson D, Wedrén S, Nordenskjöld M, Nishimura G, Grigelioniene G.
Am J Med Genet A. 2016 Jan;170A(1):266-9.

III. **Extending the phenotype of *BMPER*-related skeletal dysplasias to ischiospinal dysostosis.**
Kuchinskaya E, Grigelioniene G, Hammarsjö A, Lee HR, Högberg L, Grigelionis G, Kim OH, Nishimura G, Cho TJ.
Orphanet J Rare Dis. 2016 Jan 4;11:1.

IV. **Novel *KIAA0753* mutations extend the phenotype of skeletal ciliopathies.**
Hammarsjö A[*], Wang Z[*], Vaz R, Taylan F, Sedghi M, Girisha KM, Chitayat D, Neethukrishna K, Shannon P, Godoy R, Gowrishankar K, Lindstrand A, Nasiri J, Baktashian M, Newton PT, Guo L, Hofmeister W, Pettersson M, Chagin AS, Nishimura G, Yan L, Matsumoto N, Nordgren A, Miyake N, Grigelioniene G[#], Ikegawa S[#].
Sci Rep. 2017 Nov 14;7(1):15585.

V. **Increased yield of genetic diagnoses in skeletal ciliopathies using massively parallel sequencing, structural variant and RNA analyses.**
Hammarsjö A, Pettersson M, Chitayat D, Handa A, Taylan F, Eisfeldt J, Lagerstedt-Robinson K, Nilsson D, Girisha KM, Conner P, Papadogiannakis N, Anderlid BM, Basel D, Furuya S, Ohashi H, Nayak SS, Horemuzova E, Millar K, Godoy R, Lin A, Fukuda N, Hyodo H, Korņejeva L, Shimizu K, Beleza-Meireles A, Chung B, Wirta V, Stranneheim H, Magnusson M, Wedell A, Nordgren A, Nishimura G, Lindstrand A[#], Grigelioniene G[#].
Submitted manuscript (2018).

[*] shared first authorship
[#] shared senior authorship

RELATED PUBLICATIONS

VI. **Expanding the Clinical Spectrum of Phenotypes Caused by Pathogenic Variants in *PLOD2*.**
Leal GF, Nishimura G, Voss U, Bertola DR, Åström E, Svensson J, Yamamoto GL, Hammarsjö A, Horemuzova E, Papadiogannakis N, Iwarsson E, Grigelioniene G, Tham E.
J Bone Miner Res. 2018 Apr;33(4):753-760.

VII. **Genotype-Phenotype Correlations of *PLOD2* Skeletal Dysplasias Using Structural Information.**
Tham E, Grigelionis G, Hammarsjö A, Grigelioniene G.
J Bone Miner Res. 2018 Jul;33(7):1377-1378.

VIII. **Autosomal recessive mutations in the *COL2A1* gene cause severe spondyloepiphyseal dysplasia.**
Tham E, Nishimura G, Geiberger S, Horemuzova E, Nilsson D, Lindstrand A, Hammarsjö A, Armenio M, Mäkitie O, Zabel B, Nordgren A, Nordenskjöld M, Grigelioniene G.
Clin Genet. 2015 May;87(5):496-8.

CONTENTS

LIST OF ABBREVIATIONS

aCGH	Array Comparative Genomic Hybridization
ACH	Achondroplasia
ACMG	American College of Medical Genetics
BMD	Bone Mineral Density
BMP	Bone Morphogenetic Protein
CDG	Congenital Disorder of Glycosylation
CDT	Carbohydrate-Deficient Transferrin
CHH	Cartilage Hair Hypoplasia
CNV	Copy Number Variant
DECIPHER	Database of Chromosomal Imbalances and Phenotype in Humans Using Ensembl Resources
DMR	Differentially Methylated Region
DNA	Deoxyribonucleic Acid
ER	Endoplasmic Reticulum
ESE/ESS	Exonic Splicing Enhancer/Silencer
EURORDIS	European Organization of Rare Disorders
ExAC	Exome Aggregation Consortium
FGF	Fibroblast Growth Factor
GAG	Glycosaminoglycan
GIKANIS	Gillessen-Kaesbach-Nishimura Syndrome
GnomAD	Genome Aggregation Database
GOF	Gain-of-function
GRCh37	Genome Reference Consortium Human Build 37
HGMD	Human Gene Mutation Database
HGP	The Human Genome Project
IFT	Intraflagellar Transport
IHH	Indian Hedgehog
ISD	Ischiospinal dysostosis
ISE/ISS	Intronic Splicing Enhancer/Silencer
JBTS	Joubert Syndrome
LLO	Lipid-Linked Oligosaccharide
lncRNA	Long non-coding RNA
LOF	Loss-of-function
LOH	Loss Of Heterozygosity
MAF	Minor Allele Frequency

miRNA	microRNA
MLPA	Multiplex Ligation Probe Amplification
MPS	Massively Parallel Sequencing
MRI	Magnetic Resonance Imaging
MSC	Mesenchymal Stem Cells
NFSD	Swedish National Agency for Rare Diseases (Nationella funktionen sällsynta diagnoser)
NMD	Nonsense Mediated Decay
OFD	Orofaciodigital Syndrome
OMIM	Online Mendelian Inheritance in Man
PCR	Polymerase Chain Reaction
PGD	Preimplantation Genetic Diagnosis
pLI	Probability of gene being LOF Intolerant
PTV	Protein Truncating Variant
RNA	Ribonucleic Acid
SEDSTN	Spondyloepiphyseal Dysplasia type Stanescu
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variation
SOX	SRY-related HMG box
SRTD	Short-Rib Thoracic Dysplasia
SV	Structural Variant
TAD	Topological Associated Domain
UPD	Uniparental Disomy
VUS	Variant of Unknown Significance
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

1 INTRODUCTION

1.1 SKELETAL DYSPLASIAS

Skeletal dysplasias are a heterogeneous group of disorders primarily affecting the bone and connective tissues. The development and growth of the skeleton is tightly coordinated by the genetic program and abnormalities in skeletal signalling pathways lead to profound consequences for bone development. There are approximately 1 000 genes important for skeletal development and mutations in many of them lead to congenital skeletal disorders. Each single clinical entity of skeletal dysplasias is rare and altogether there are about 450 conditions,[1] many of them have not been well characterized yet and the molecular pathways for many of the reported conditions have not been explored in detail.

Even though there is no general consensus on the definition of a rare disease, according to the European organization of rare disorders (EURORDIS; www.eurordis.org), a rare disorder is one that affects less than 1 in 2 000 individuals. In Sweden though, a rare disorder is defined as 1 in 10 000 (NFSD; www.nfsd.se). This means that a rare disorder may affect only a small percentage of patients in the world, while others affects as many as 200 000. Congenital skeletal disorders occur approximately in two to three in 10 000 individuals.[2-4] Two common skeletal dysplasias are Leri-Weill dyschondrosteosis (LWD [MIM:127300]) affecting approximately 1:2 000 individuals and achondroplasia (ACH [MIM:100800]) with an incidence around 1:20 000. However, the prevalence is very different among skeletal dysplasias, with some of the them being extremely rare and so far only reported in a few individuals in the literature, for example al Gazali skeletal dysplasia[5] and Saul-Wilson syndrome.[6]

Genetic skeletal disorders include dysplasias, dysostoses, and osteolyses, and the distinction between these conditions is essential for accurate diagnosis. Skeletal dysplasias comprise a diverse group of diseases most often characterized by short stature due to abnormal growth of bones and cartilage organization or due to mineralization defects. The genes involved in skeletal dysplasias are thought to be constantly expressed from early embryonic stages throughout adulthood. In contrast, dysostoses are defects localized to a single skeletal component (or a few of them in combination), due to deficient regulatory factors only temporarily expressed during embryogenesis (within the first eight weeks of embryonic development). Skeletal dysplasias and dysostoses overlap, but skeletal dysplasias are more generic and usually progress as a result of the persistent expression of an affected gene through skeletal development and longitudinal bone growth. In osteolyses, on the other hand, the bone formation is normal at first, but then a process of progressive destruction or disappearance of bone tissue starts.[7]

Skeletal dysplasias should per definition be distinguished from skeletal syndromes, since dysplasias occur in isolation (non-syndromic). In skeletal syndromes bone and cartilage abnormalities are associated with a broad spectrum of symptoms from other organs, such as hearing impairment, internal malformations or dysfunction of the immune system. However,

despite the above theoretical definitions, in clinical practice, the terms “skeletal dysplasia” or “skeletal disorder” are commonly used and these terms will be used in this thesis.

1.1.1 Clinical presentation

Careful clinical examination and detailed radiographic evaluation of the patients with congenital skeletal dysplasias are essential for correct diagnosis. Clinical examination includes morphological top-to-toe examination, skeletal survey, and often other imaging techniques when structural anomalies are suspected (such as ultrasound or MRI). Specific radiographic signs in combination with clinical findings, such as dysmorphic features, internal malformations or biochemical abnormalities usually lead to a diagnostic hypothesis, as exemplified in **Figure 1**.

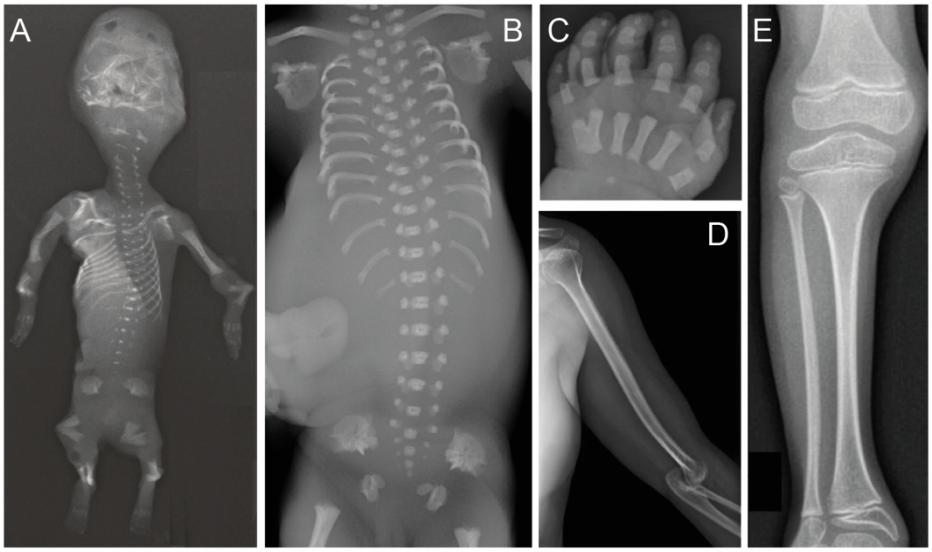

Figure 1. Examples of radiographic signs in patients with skeletal dysplasias. A. Fetus at GA 15 weeks (missed abortion) with kyphomelic dysplasia, poor bone mineralization, deformed skull, trunk, and short angulated bones. **B**. Fetus with short-rib thoracic dysplasia at 21 weeks GA (termination of pregnancy) with short ribs, short iliac bones, and trident pelvis. **C.** Hand shows brachydactyly and postaxial polydactyly. **D**. Radiographs of upper arm of an adult patient with De Hauwere syndrome, note gracile humerus, and elbow dislocation. **E**. Lower extremity radiograph of a 5-year-old child with Saul-Wilson syndrome with slender tubular bones with metaphyseal flaring and large epiphyses. GA, gestational age.

Historically, classification of skeletal dysplasias was based on phenotypic and radiographic appearance. Upon the discovery of the genetic causes some of them were divided into groups according to the underlying genetic defects. Currently, some groups in the nosology of congenital skeletal disorders are based on genetic background, while others are named according to the anatomical position of the most predominant abnormalities seen in the skeleton[1]. For example, in spondylometaphyseal dysplasias the main radiographic abnormalities are found in the spine (*spondylo*) and the *metaphyses* of the long bones, acromesomelic dysplasias have short bones in the hands and feet (*acromelic*), and short middle parts of the limbs (*mesomelic*), while cleidocranial dysplasia is named after the abnormal clavicles (*cleido*) and *cranium*.

Within the same nosology group there may be a significant overlap in the clinical presentation between different skeletal dysplasias.[1] Genetic tests are therefore useful for differentiation,

because phenotypically overlapping conditions are often caused by pathogenic variants in genes coding for different molecules involved in the same cellular signalling pathway.

The improvement in genetic diagnostics of skeletal disorders leads to a deeper knowledge about pathogenesis and allows accurate genetic counselling with information about recurrence risk in the affected family. Additionally, it may provide insights of the natural course of the disorder (learning from previously reported affected individuals) and potential complications that might be progressive or fatal if untreated. Family history may sometimes give clues to whether the skeletal dysplasia is caused by a *de novo* mutation or if it's recessively inherited, but a precise molecular diagnosis is a prerequisite for the availability of future prenatal diagnosis and preimplantation genetic diagnosis (PGD).

1.2 GENETICS

In 1990, the road map to the human genome – the human genome project (HGP) – was initialized. It was an international research effort to sequence and localize all human genes, including protein domains and alternative isoforms. In 2001, the first draft of the human genome was released[8,9] and in 2003 the project was announced completed.[10] The main conclusion from this work was that the human genome contained much fewer genes than expected, with predictions pointing to between 60 000-120 000 genes. Today it's estimated that the human genome contains about 3 billion nucleotides and approximately 20 000 protein coding genes. However, the current assembly of the human genome (GRCh37/hg19) still has gaps and certain genomic features, such as repetitive regions and pseudogenes, that remain problematic to study.[10,11]

Organism complexity is not dependent on features such as genome size or number of genes in a species.[12] However, alternative splicing, the mechanism when one gene produces more than one protein isoform and diversifies protein function is linked to human complexity. Most of the human genes (with more than one exon) undergo alternative splicing.[13] Another important difference in organism complexity is the non-coding DNA, the functional elements that only produce RNA molecules. They are transcriptional regulating units, like microRNAs (miRNA) or long non-coding RNAs (lncRNA), that act as regulators, gene silencers or activators.[14-16] The different range of post-transcriptional regulation, transcripts, and proteins in the cells and tissues – the transcriptome and proteome – are the links that decipher how our DNA leads a cell to normal development and function, or to a specific disease phenotype.

In the last decade, there has been a great increase in the number of known disease genes and chromosomal aberrations in genetic diseases. The genetic diseases identified to date, are listed in the Mendelian Inheritance in Man (MIM) and accessible in the OMIM database, where over 15 500 genes and 7 800 clinical entities with known genetic causes are catalogued.[17] Due to the diversity and phenotypic overlap among rare diseases, it is a challenge to diagnose patients with rare disorders solely based on phenotype. For extremely rare conditions, it is possible that a physician has never met a patient with that specific diagnosis before. Furthermore, the correct diagnosis may be difficult to pinpoint because of phenotypic variability. Today, there are many ongoing efforts (regional, national, and

international) to organize centres for patients to increase knowledge, experience, and research collaborations on rare diseases.

The bioinformatic tools used to draft the human genome were also used to sequence the genomes of other organisms, such as mice and zebrafish. The results have been made publicly available allowing human genetic sequences to be compared to sequences of other organisms. Comparative genomics is of great importance since there are lots of conserved genes between organisms (i.e. orthologs) and if we can explain the function of a gene in a model organism, we might be able to deduct its function in humans leading to a broader understanding at the molecular level. Moreover, comparable genomics of organisms closely related to humans can give information about important features of DNA, like exons and regulatory elements (since they are more conserved).

1.2.1 Genetic loci for skeletal dysplasias

Most of the skeletal dysplasias are monogenic disorders and follow the pattern of monogenic inheritance: autosomal dominant, autosomal recessive, X-linked dominant, and X-linked recessive. In cases where the same phenotype can be caused by mutations in several genes or when the genetic aetiology is unknown, massively parallel sequencing (MPS) technologies using whole exome or whole genome sequencing (WES/WGS) is the golden standard to quickly achieve molecular diagnosis. Since WES and WGS provide many rare variants, correlation with the clinical and radiographic findings is essential for correct variant interpretation in skeletal dysplasias.

The classification meetings for congenital skeletal disorders are held regularly in association to the International Skeletal Dysplasia Network meeting by a group of experts, who summarize the latest classification in a scientific report, updated every 4-5 year. In the last revision of the nosology and classification of genetic skeletal disorders from 2015 there are 436 conditions, which are divided into 42 groups and caused by mutations in 364 genes.[18] The number of genes has increased dramatically since the previous revisions, from around 80 in the 2001 classification[19] to 226 reported disease-causing genes in the 2010 nosology report.[20] The nosology could be seen as a "master list" of skeletal dysplasias or an overview of the different nosology groups, including genetic information. A clinical entity is placed in the nosology if it has been described in the literature and is included in a group because of similar pattern of skeletal involvement and/or a common genetic cause or molecular pathway. Before a gene is associated with a specific condition in the nosology, the genetic basis is proven in multiple patients and/or pathogenicity is confirmed with functional analyses.

There are thousands of genes in which pathogenic variants are reported to cause monogenic diseases, but the genetic basis is still unknown for a significant number of rare diseases, including many skeletal dysplasias. This could be due to disease-causing genes that are not identified yet or genetic variants in known disease genes difficult to interpret or detect. Using current bioinformatic pipelines regulatory regions or other non-coding variants, such as synonymous variants and deep intronic variants that affect splicing, or middle-sized structural rearrangements are hard to find. Another scenario is digenic or oligogenic inheritance, when variants of two or more disease-causing loci segregate with disease. As an

example, triallelic inheritance have been reported by Katsanis *et al.* in Bardet-Biedl syndrome (BBS [MIM:209900]), which is when a combination of three pathogenic variants in two different genes are necessary to cause the disorder.[21] There are also a few studies supporting digenic inheritance in BBS[22] and it has been reported in skeletal ciliopathies once.[23] On the other hand, a combination of all deleterious variants in an individual, the mutational load, might also explain incomplete penetrance, severity, and onset of disease.[24]

1.2.1.1 Single gene defects

To date, there are more than 450 genes associated with skeletal dysplasias, commonly included in smaller gene panels used for diagnosis. Using WGS or WES with panel-based filtering allows analysis of variants in the included genes. Over time, the *in silico* gene panels have been extended, with inclusion of new disease-causing genes and variants as they are reported. Furthermore, if a panel analysis fails to identify disease-causing variants, DNA samples of family members may be added for a trio/quattro molecular analysis. This approach may be necessary to achieve molecular diagnosis when the variants are in novel genes, not yet associated with a phenotype.

Despite the genetic heterogeneity of skeletal dysplasias there are some with easy clinically recognizable phenotypes, such as cartilage hair hypoplasia, dyschondrosteosis, achondroplasia, pseudoachondroplasia, and hypochondroplasia. In these cases, the diagnosis is based on the clinical and radiological findings, and a single gene analysis. Molecular analysis is performed using conventional Sanger sequencing if the genes are small or there is a mutation hot-spot in the gene. For example achondroplasia is caused by a p.Gly380Arg substitution in the *FGFR3* gene in 98% of the cases[25] and the Finnish founder mutation in *RMRP* (c.70A>G) is very common in patients with cartilage hair hypoplasia (CHH [MIM:250250]) in northern Europe.[26]

Even though some skeletal dysplasias are easy to diagnose, others are more variable with multiple phenotypes and genetic overlap, making Sanger sequencing laborious. For these entities, whole genome sequencing is more effective, allowing evaluation of both sequence and gene dose abnormalities. For example, at least 15 phenotypically different clinical skeletal entities caused by mutations in *COL2A1* have been described so far. However, for most of them, radiographic features such as platyspondyly, “pear-like” vertebral bodies, and hypoplasia of the pubic bones are key radiographic features. Type II collagen is the major collagen synthesized by chondrocytes and the clinical variability and phenotypic overlap in these disorders are commonly observed, even within the same family.[27]

In this thesis I have examined not only pure skeletal dysplasias but also skeletal syndromes, with internal malformations caused by metabolic and cilia abnormalities, therefore there is a need to introduce the congenital disorders of glycosylation (CDG) and the skeletal ciliopathies in the next two sections.

Metabolic disorders with skeletal involvement

Several metabolic diseases with a skeletal phenotype, have been described to date, such as mucopolysaccharidoses [MIM:607014] and congenital disorders of glycosylation (CDG). Mucopolysaccharidoses are a group of lysosomal storage disorders with a defect regulation of glycosaminoglycan (GAG) metabolism, that results in accumulation of GAGs in the cell and lead to a range of clinical features, including skeletal abnormalities with severe handicap.[36] There are at least ten genes known to cause different forms of mucopolysaccharidoses. CDG is another group of metabolic disorders with either an autosomal recessive or X-linked inheritance and a vast number of clinical features. CDG patients show a wide range of biological abnormalities in protein glycosylation.[37] Glycosylation is a complex post-translational modification, attaching sugar to specific amino acids in many proteins (**Figure 2**), important in protein folding and transport.[38]

In the majority of the cases, the disorder begins early in life and symptoms range from severe developmental delay and hypotonia with many organs involved, to milder symptoms. To date, there are more than 70 entities of the N-linked CDGs, but many of them have only been described a few times, limiting the phenotype-genotype correlation. The nomenclature for CDGs involve the gene name followed by CDG, for example ALG9-CDG. The consequences of an impaired glycosylation in skeletal dysplasias are not well-defined, but with more patients being reported with skeletal phenotypes in CDG, this diagnosis should be considered for unknown multisystem diseases, with symptoms including the skeleton.[37]

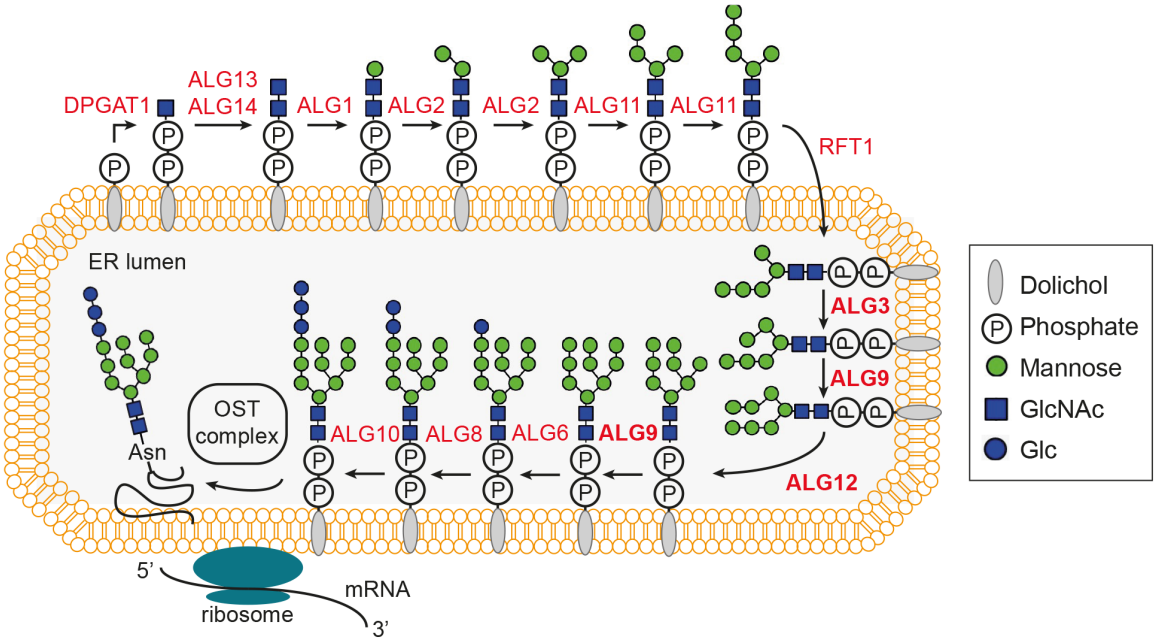

Figure 2. Simplified scheme of the N-glycosylation pathway in the ER. The protein and LLO are generated separately and after subsequent enzymatic steps LLO is transferred into the luminal side of ER. Genes (in red letters) have been implicated in CDG. Severe skeletal changes have been reported in patients with pathogenic variants in *ALG3*, *ALG9,* and *ALG12*, which add mannose to the growing LLO. Mild skeletal changes (brachydactyly) have been reported in patients with disease-causing variants in *ALG6* and *ALG8*, which add the first and second glucose. (Patients with ALG6-CDG and ALG8-CDG are also associated with many other features). In the end, the OST complex transfers LLO from the dolichol and attaches it to the asparagine residue on the protein in the ER lumen. Asn, asparagine; ER, endoplasmic reticulum; Glc, glucose; GlcNAc, N-acetylglucosamine; LLO, lipid-linked oligosaccharide; OST, oligosaccharyltransferase. [This figure is modified from Ng BG et al.[39] with permission from *Elsevier*]

Ciliopathies with major skeletal involvement

Ciliopathies is a very broad group of genetic disorders caused by defective cilia functions. Within the skeletal ciliopathies there are multiple disease phenotypes with a phenotypic overlap and multiple disease-causing genes.[28,29] Almost all cells in the body have cilia, which play an important role in organogenesis, cell migration, and signalling, including the skeleton formation. Ciliary proteins are synthesized in the cytoplasm and transported into the cilia in a process called intraflagellar transport (IFT). To date, pathogenic variants in at least 30 genes are known to cause skeletal ciliopathy phenotypes and all of them encode proteins either localized to the IFT machinery or is important for cilia structure and function.[30] The cilia function as a signalling centre and are needed to maintain the cell organization in the growth plate and disruption of the cilia is thought to affect the Indian hedgehog (IHH) signalling pathway in skeletal ciliopathies.[31]

The skeletal ciliopathies are classified according to their radiographic features and the most common clinical entities are short-rib polydactyly syndromes (SRPS) and asphyxiating thoracic dystrophies (ATD). In OMIM, they are grouped into a broader group of short-rib thoracic dysplasias (SRTD [MIM:208500]), Ellis-van Creveld syndrome (EVC [MIM:225500]), and Sensenbrenner syndrome (also known as cranioectodermal dysplasia; CED [MIM:218330]). The main characteristic skeletal features include narrow thorax, short tubular bones, metaphyseal flaring, hypoplastic lower part of the pelvis, and short hands with or without polydactyly.

Deleterious variants in genes coding for proteins in the cilia compartments have similar effects on cilia function, leading to phenotypically overlapping conditions, likewise mutations in the same gene can lead to other ciliopathy syndromes. For example, mutations in *C2CD3* have been reported in both SRTD[32] and Joubert syndrome (JBTS [MIM:213300]),[33] both in combination with orofaciodigital syndrome (OFD [MIM:617127]) and mutations in *CEP120* can lead to complex ciliopathy phenotypes[34] and SRTD.[35]

1.2.1.2 Chromosomal aberrations

Larger chromosomal aberrations include deletions, tandem duplications, inversions, insertions, and translocations (**Figure 3**), comprising of only a few exons of a disease-causing gene to a whole chromosome. Deletions and duplications are copy number variants (CNV) with loss or gain of genetic material, compared to the reference genome. The size cut off for a CNV stretches >50 base pairs within the genome, although the size is arbitrary and depends on technology limitations.[40] Most of the skeletal dysplasias are monogenic disorders caused by small changes and only a few are caused by recurrent rearrangements (larger duplications or deletions). *SHOX* haploinsufficiency, including recurrent deletions and CNVs in flanking regulatory elements is one of the well-known examples.[41,42] Any chromosomal aberration involving a gene important for skeletal development will most probably lead to skeletal abnormalities and additional features depending on whether other genes are affected by haploinsufficiency. For example, such as in the 1q24q25 deletion syndrome, comprising of at least thirteen genes, where affected individuals have learning disabilities and distinctive facial characteristics, in addition to short stature and brachydactyly.[43]

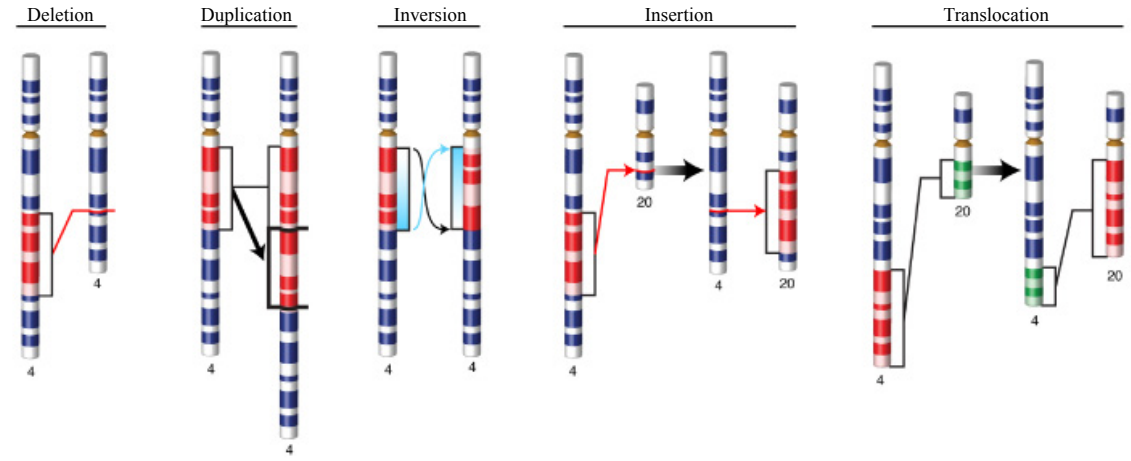

Figure 3. Schematic overview of large structural variants showing chromosomal abnormalities on chromosome 4 and 20. Even though inversions, insertions, and balanced translocations do not lead to loss (deletion) or gain (duplication) of genetic material, these events can result in gene disruption and phenotypic consequences. [Courtesy of National Human Genome Research Institute and the Smithsonian National Museum of Natural History]

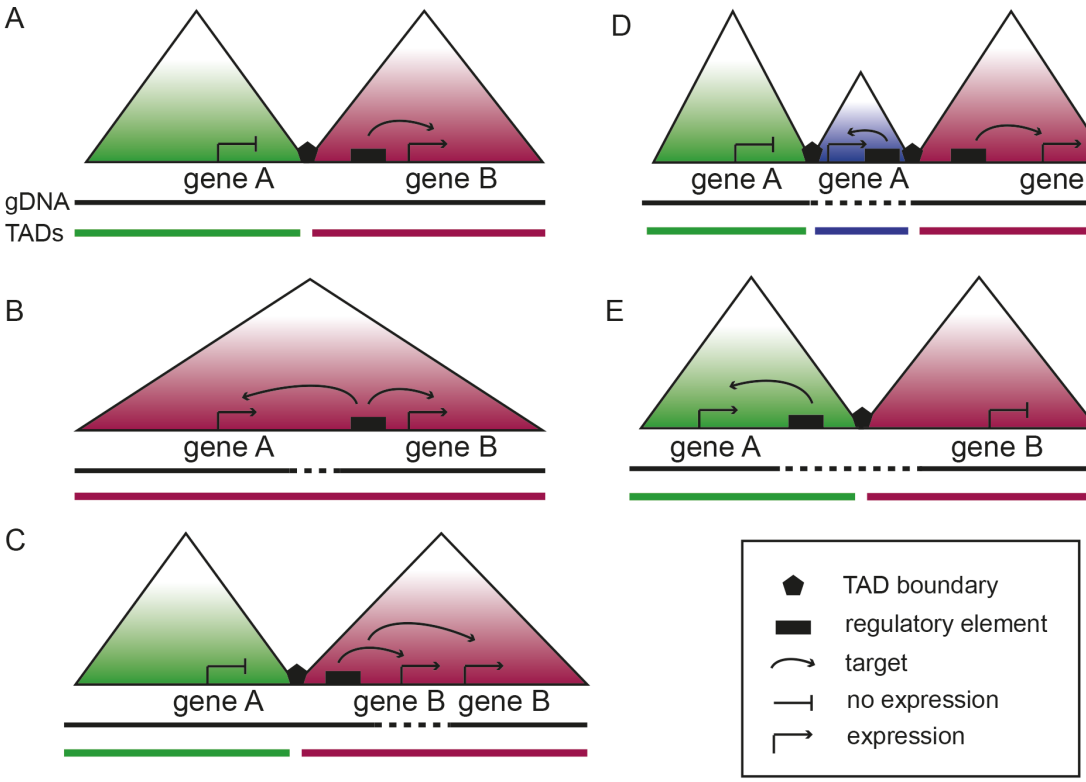

Figure 4. Gene expression regulation within TAD domains and effect of structural variants. A. Healthy genome with two TADs that separate inactive gene A from the regulatory element and expressed gene B. **B.** Deletion of a TAD boundary causes a fusion of the two TADs and the regulatory element can also activate gene A. **C.** Duplication within the TAD (intra-TAD) making two copies of gene B. **D.** Duplication involving the TAD boundary between two TADs (inter-TAD) and a regulatory element from neighbouring TAD activates gene A in the new TAD (neo-TAD). **E.** Inversion containing the TAD boundary and regulatory element enable activation of gene A. The TAD containing gene B lost its regulatory element and gene B is repressed. Dashed line indicates the genomic structural variant. TAD, topologically associated domain.

It is known that structural variants or rearrangements outside the coding region can impact gene expression if they affect promoters or enhancers of disease genes. In recent years, studies using conformation capture technologies show the 3D-organization of the genome and describe topologically associating domains (TADs) to be responsible. TADs are important to promote the contact within a domain (promoter and/or enhancer will physically be close to the gene), but there is very little contact between different TADs. It is now clear that altering a TAD boundary can have impact on gene expression, because regulatory elements within TADs can affect gene expression and in some cases result in disease (**Figure 4**).[44,45] In mouse, the Shh region have been extensively studied for how TADs are responsible for the anterior-posterior positioning of the limb bud. It has been shown that disruption of the TAD boundaries alter the communication between Shh and its regulatory element and cause truncation of the limbs, while deletions within the TAD have no impact of limb formation.[46] In humans, preaxial polydactyly have been shown in individuals with duplication spanning the regulatory element of *SHH*.[47]

1.2.1.3 Imprinting defects with skeletal abnormalities

Genetic imprinting means that different genes are expressed dependent on parent-of-origin of the methylated regions and it is known to be important for normal development. There are very few syndromes with skeletal symptoms associated with imprinting abnormalities. Some imprinting defects are classified as differential diagnoses to skeletal dysplasias, while others are included in the nosology of skeletal dysplasias.[1] One imprinting defect with skeletal features is the Kagami-Ogata syndrome (KOS14 [MIM:608149]) that is caused by either a maternally inherited microdeletion of a differentially methylated region (DMR) on chromosome 14q32.2 or by paternal uniparental disomy of chromosome 14 (UPD14). Opposite to paternal UPD14, maternal UPD14 is known as Temple syndrome (TS14 [MIM:616222]) with short stature, intellectual disability, and early onset of puberty.[48] A differential diagnosis to TS14 is Silver Russel syndrome (SRS [MIM:180860]) caused mainly by loss of methylation of chromosome 11p15 or maternal UPD7. Other examples include pseudohypoparathyroidism Ia (PHP1A) and pseudopseudohypoparathyroidism (PPHP), both presenting with a phenotype of Albright hereditary osteodystrophy with or without hormone resistance respectively, and both are caused by mutations in the *GNAS* locus.[49]

1.2.1.4 Non-coding RNAs

Even though, most of the disease-causing variants reside within a protein coding gene, some studies suggest a role for non-coding RNAs in skeletal dysplasias and syndromes.[16] CHH is a skeletal dysplasia that is relatively common in Finland, with symptoms including short stature, a high risk for cancer, and immune disorders. The responsible gene was found in 2001, *RMRP*, a gene not coding for a protein, but an RNA molecule important for mitochondrial transcription and replication.[26]

There is a large group of small non-coding RNAs, the microRNAs (miRNAs), regulating gene expression. They have shown to target a vast amount of human genes and many of the miRNAs lie within genes both in introns and exons. Many of the miRNAs regulate key

proteins in skeletal development and depending on target they regulate gene expression by selectively silencing them. One miRNA can target several hundred genes and one gene can have different binding sites for more than one miRNA, making this to a complex network of gene regulation, which is often tissue specific.[16] There have not been many cases reported with skeletal symptoms in humans that are caused by disruption of a miRNA, but one example is the deletion of a miRNA cluster (miR-17~92) that causes brachydactyly, scaphocephaly, and short stature.[50] With the increase of studies performed on miRNA, the list of miRNAs causing diseases will most likely increase in the future.

1.2.2 Interpretation of variants

The diagnostic yield for skeletal dysplasias using MPS based panels and exomes is around 60%,[51-56] but seems to depend on skeletal phenotype in the cohorts. Studies including a broad range of skeletal dysplasia patients with many genes have lower outcome, while more defined cohorts like osteogenesis imperfecta or skeletal ciliopathies have been reported with genetic diagnoses in more than 80% of the cases.[52,55] There is a great challenge in interpreting the identified variants, when a single individual may harbour millions of variants and only one or a few are disease-causing. The rest of the variants account for normal variation.

A single nucleotide variant (SNV) refers to that one nucleotide is replaced by another one, sometimes also referred to single nucleotide polymorphism (SNP). However, SNV is favoured over SNP, since a polymorphism implies that it is a common variant with a frequency of at least 1% in a population. Each individual is predicted to carry at least three million SNVs and around 1 000 CNVs, which means that the genetic variation between two individuals differ roughly about 0.1% due to SNVs and 1% due to CNVs.[57] The genetic diversity differs between populations and it have been shown that African populations have more genetic variations compared to other populations outside Africa. This seem to match the out-of-Africa hypothesis which resulted in a population bottleneck and loss of genetic diversity.[58]

To improve the variant analysis, the concept of minor allele frequency (MAF) was introduced. It defines the frequency of a variant (the less common allele) at a certain position in the genome in a population. The allele frequency of a SNV differ between populations and ethnical groups as well as for certain diseases. This is important to have in mind when interpreting a variant, because a variant that is rare in one population could be common in another and should not be considered as pathogenic, and the incidence of a disease is dependent on MAF. The 1000G project, SweGen, ExAC, and gnomAD consortium have made databases available with more than 120 000 exomes and genomes,[59-61] with the invaluable landscape of human genetic variation needed to filter MPS data. The SweGen data set comprise of 1 000 individuals from the Swedish population and is a valuable source for local genetic variation and since many of our samples are run on the same platform, detecting sequencing errors as possible pathogenic variants are minimized.[60]

There are different types of SNVs in the DNA code: missense, silent, frameshift, and protein-truncating variants (PTVs). PTVs include introduction of a stop codon (nonsense) and splice variants caused by SNVs. The missense variants are the most common and could easily be

checked for conservation across species using SIFT (Sorts Intolerant From Tolerant)[62], PolyPhen2 (Polymorphism Phenotyping version 2),[63] and GERP++,[64] but for nonsense, synonymous, and PTVs these prediction tools are poor. Combined Annotation Dependent Depletion (CADD) is another tool that measures the deleteriousness for all variants across the genome and integrates multiple annotation tools into a C-score for interpretation.[65] This value is useful in prioritizing all disease-causing variants.

The deleterious alleles tend to be rare and population allele frequency from ExAC have been used in attempt to identify genes with a high probability of loss-of-function intolerance metric score (pLI >0.9).[59] This is applicable for PTVs and genes that cause dominant disorders, while the constraint metric gives Z-scores for CNVs, synonymous, and missense variants. Positive Z-scores suggest an intolerance to variation in the gene and if the Z-score is >3.09 there is less observed variants than expected, while negative Z-scores have more variations than expected and are tolerant to variation.[59] Interpretation of variants can therefore be guided by the use of constraint scores.

Many variants are extremely rare and have just recently been introduced in the coding sequence, thus only shared by close relatives or cluster in a population. *De novo* variants occur in each generation with an incidence of 1.0-1.8 every 10^8 nucleotides per generation. Each individual carries around 50 to 70 SNVs and less than ten smaller indels (<50 bp) in the genome, that neither of their parents have.[66,67] One or two of these SNVs affects the coding region, but is not necessarily causing disease. Therefore, the pathogenicity of a variant cannot solely be based on if the variant is *de novo* or not. Another limitation is the functional impact of the variant that can be hard to predict. One variant might lead to a complete or partial loss-of-function (LOF, so called inactivating variants), reducing the proteins normal function. Other variants may lead to activation of the protein (so called gain-of-function variants, GOF), that either lead to a stronger activation of the protein function or a totally new function (neomorph). The distinction between these different forms of variants are usually only clear when performing functional studies.

In summary, after filtering for a variant using gene, amino acid, and nucleotide conservation, constraint metrics, allele frequency, and inheritance pattern, as well as previous reports in disease-variant databases (i.e. HGMD, ClinVar, DECIPHER) the amount of deleterious variants will decrease significantly. Hopefully, there are only a few left to interpret further for clinical relevance and disease mechanism. Using the American College of Medical Genetics and Genomics (ACMG) variant classification system the remaining variants may be placed into one of five categories (benign, likely benign, variants of unknown significance (VUS), likely pathogenic, and pathogenic).[68] After whole exome or genome sequencing without filtering for a gene panel many variants will fall into the VUS category. Finally, it is important to remember that variants that have been reported previously as pathogenic, may actually be benign and *vice versa*.[69,70]

Currently, due to the insufficient understanding of non-coding variants, their interpretation is not as accurate as the coding variants. In fact, despite the knowledge available on non-coding regions, most variants that are considered are located within protein-coding genes.[71]

1.2.2.1 Splice variants

Pre-mRNA splicing is mediated by the spliceosome, a large complex of proteins, that recognizes the canonical splice sites at the acceptor and donor splice site, the branchpoint, the polypyrimidine tract, and other recognition sites (**Figure 5A**).[72] The spliceosome consists of five small nuclear ribonucleoproteins (snRPs) and is dependent on hundreds of splicing factors that are usually expressed in cell-type specific fashion. During transcription, an organized series of events lead to the cleavage of phosphodiester bonds at the exon-intron boundaries and the exons are ligated, whereas the introns are removed. The awareness that disease-causing variants may affect regulatory elements in splicing have led to a different approach in evaluating possible splice variants, though they are difficult to interpret[73] and most likely underestimated. Correct interpretation of splice variants will most likely increase diagnostic yields.

Several studies show that up to 10% of reported missense and synonymous variants actually affect pre-mRNA splicing.[72,74] SNVs can be located within and outside of the coding exons of a gene and still can have an effect on splicing and protein expression. The spliceosome binds to recognition sites to regulate splicing, which can be both exonic and intronic. These regulatory elements can work as either splicing silencers or enhancers.[75] Disease-causing splice variants most often occur at the exon-intron boundary, ±2 base pairs from the canonical splice site and usually lead to exon skipping, but there are also deep intronic splice variants that lead to exon inclusion.[76] Other variants may introduce a cryptic splice site in close proximity to the existing splice site, causing partial retention of intronic sequence within the transcript, deletion of the exonic sequence upstream of donor splice site or retention of the whole intron in the mRNA (**Figure 5B**).

Canonical and non-canonical splice variants can lead to both LOF and GOF mutations. The functional effects of these mutations depend on which gene and position of the variants and if they lead to in-frame or out-of-frame translation, as well as if nonsense mediated decay (NMD) occurs or not.[77]

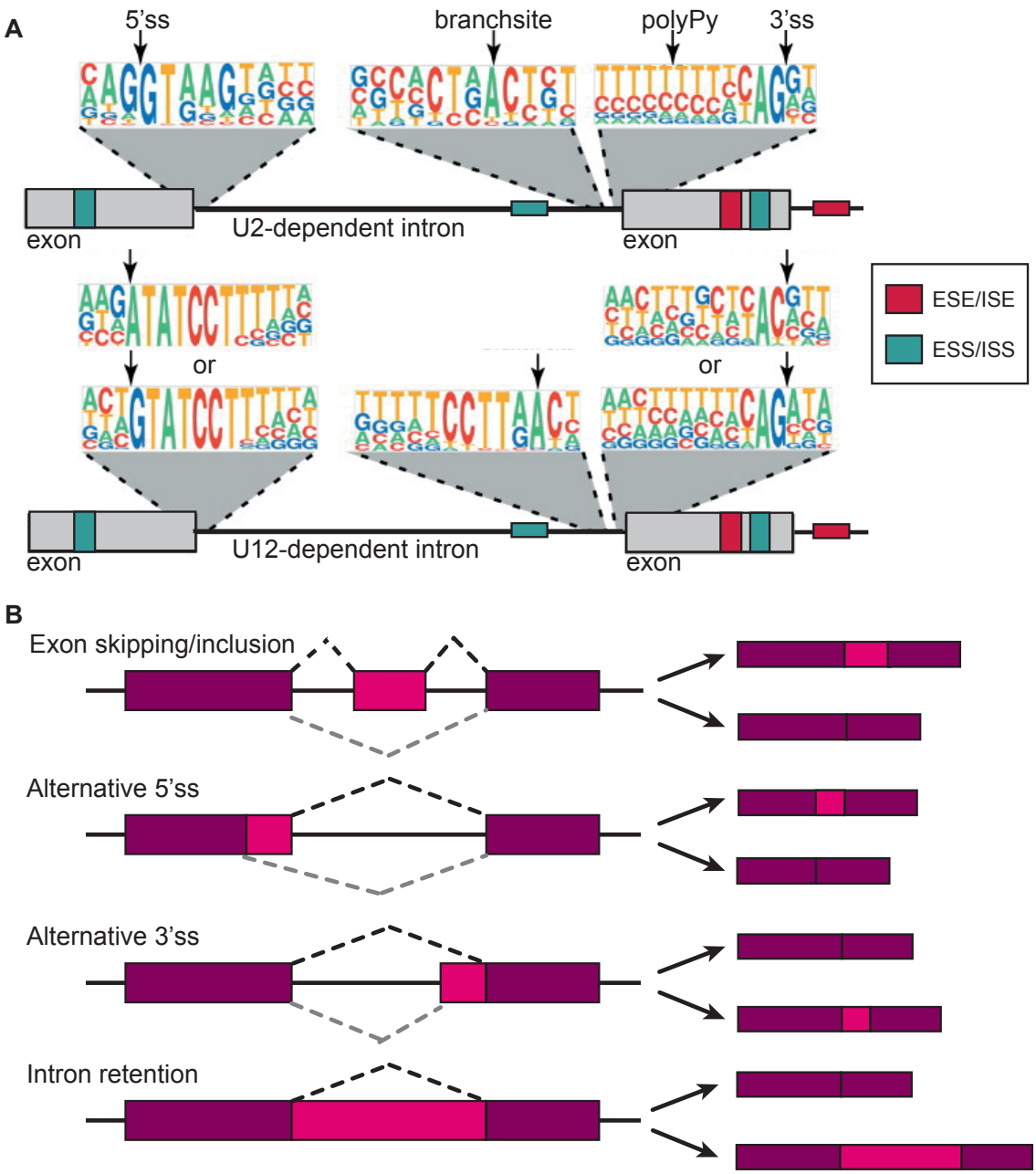

Figure 5. Classical splicing signals for U2 and U12-dependent human introns with regulatory elements (A) and modes of alternative splicing (B). **A.** Consensus sequences of splice recognition sites, where the letter size corresponds to its frequency at each position. The U2- and U12-dependent introns differ by their dinucleotide composition at the canonical splice site and branchsite motif. In many introns upstream of the 3' splice site there is a polypyrimidine tract (usually represented by cytosine or thymine) and a branchsite 20-50 base pairs upstream of the polyPy. Additional regulatory elements in the exons or introns are recognized by the splicing machinery for correct splicing to occur (exemplified as red or green bars). **B.** Four common modes of alternative splicing that can occur and lead to different outcomes. Dashed lines indicate splicing options. ESE, exonic splicing enhancer; ESS, exonic splicing silencer; ISE, intronic splicing enhancer; ISS, intronic splicing silencer; polyPy, polypyrimidine tract; ss, splice site. [Permission obtained from *Elsevier* to reuse parts of figure 1 from Padgett et al.[72]]

1.2.3 Reverse phenotyping

Clinical diagnosis of a skeletal dysplasia is many times established by an experienced radiologist, geneticists or paediatrician based on pattern recognition. However, if a patient has an atypical phenotype, reverse phenotyping can be used, analysing the genetic data first and correlating the clinical findings next. In some cases, the clinical diagnosis can be revised after genetic testing or it may lead to the identification of a novel gene behind the atypical phenotypes. There is a small proportion of patients who are affected by more than one genetic condition,[78,79] this is mainly true for consanguineous families[80] and it should be addressed when expanding the phenotypic spectrum of a known condition, including skeletal dysplasias.

1.2.4 Phenocopies of skeletal dysplasias

There are a number of factors that may lead to conditions, mimicking skeletal dysplasias, that are not caused by genetic factors. Internal and external factors, including nutritional deficiencies (for example insufficiency of vitamin D), maternal infections (for example Rubella and Zika virus), unrelated pharmacological treatments that influence skeletal development (such as prenatal exposure to warfarin), and maternal autoimmune diseases with autoantibodies crossing the placenta during pregnancy (such as systemic lupus erythematosus, mixed connective tissue disorder, and Sjögren syndrome).[81] Therefore an adequate disease history is necessary in order to obtain the correct diagnosis.

1.3 NORMAL AND ABNORMAL DEVELOPMENT OF THE SKELETON

There are 206 bones in the human skeleton that function as a structural framework for the body and protect the inner organs. The skeleton also facilitates movement and has a metabolic role as a supply of minerals, cytokines, and growth factors needed for its modelling. The skeleton is formed at the embryonic stage and the growth of bones continues to the mid-twenties. In order to retain the skeleton, the bone mass is preserved in adult life through the action of bone formation and bone resorption.[82] Different signalling pathways define how the bone forms and how it is shaped into the right size and orientation.

1.3.1 Bone formation

The process of bone formation in the embryo is initiated by mesenchymal stem cells (MSC, mesoderm and neural crest cells), following condensation and differentiation. There are two main processes of how bone is built; endochondral and intramembranous ossification. The long tubular bones are built through the process of endochondral ossification, whereas the flat bones are built by intramembranous ossification.[83] Numerous genes control bone formation and metabolism, and recognition of signalling pathways responsible for disease provides us with insight on how the molecular mechanisms work in normal development.

1.3.1.1 Endochondral ossification

In endochondral ossification, in the early embryo, the differentiation of the MSC into chondrocytes is dependent on SOX transcription factors and bone morphogenetic protein

(BMP) receptors. SOX9 is crucial for this process shown by mutations in *SOX9* resulting in campomelic dysplasia (CMDP [MIM:608160]), often a semilethal disease with bowed femora and small thoracic cage.[84]

In cartilaginous bones, the anlagen is formed from condensed MSC and ossification begins in the central part of it in the primary ossification centre, which gradually forms the diaphysis. Later, at the end of the cartilaginous anlagen, secondary ossification centres develop and establish the epiphyses, which are located in the proximal and distal parts of the long bones. The cartilage layer between the primary and secondary ossification centres is called the epiphyseal growth plate (**Figure 6**).

In the growth plate, the chondrocytes grow towards the metaphysis, forming columns and are divided into four zones; resting, proliferative, hypertrophic, and mineralization zone. The resting chondrocytes serve as a supply of stem cells. When the stem cells divide, they form well-defined columns of chondrocytes, defining the proliferative zone, which synthesize collagen type II, aggrecan, and fibroblast growth factors (FGFs). In the hypertrophic zone, the cells increase their volume and produce more extracellular matrix which calcifies. The matrix is invaded with vessels and osteoblasts which will form the mature trabecular bone.[83] The perichondrium is a layer of connective tissue that surrounds the cartilage of developing bones important for growth regulation. Growth plates are fused in the end of puberty, which in case of skeletal dysplasia leads to a less specific radiographic appearance of the bone and the diagnostic signs diminish or disappear.[83,85]

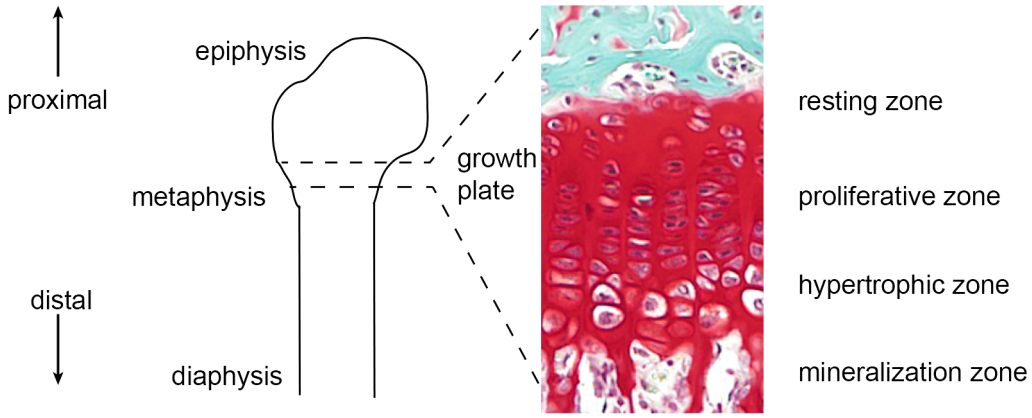

Figure 6. Schematic representation of the growing long bone. The epiphyseal growth plate from proximal tibia of mouse is stained by safranin and fast green. The growth plate can be divided into four zones; resting, proliferative, hypertrophic, and mineralization zone.

Cell proliferation and differentiation in the growth plate is a complex process and is tightly regulated by hundreds of genes, among them *SOX9*, *FGFR3, IHH, PTHLH, COL2A1, COL10A1*, and *RUNX2*.[83] Mutations in these genes cause different types of skeletal dysplasias. Indian hedgehog (IHH) plays an important role in the regulation in endochondral ossification. It is expressed in the prehypertrophic and hypertrophic chondrocytes in the growth plate and participates in a negative feedback loop involving parathyroid hormone related protein (PTHrP, encoded by the *PTHLH* gene), coordinating chondrocyte proliferation and hypertrophy. IHH induces PTHrP production in the articular end of the long

bones and PTHrP maintains the chondrocytes in proliferative state and delay the transfer to prehypertrophic chondrocytes by binding to its receptor, PTHR1.[86] In agreement with this mechanism, activating mutations in *PTHR1* cause autosomal dominant metaphyseal chondrodysplasia type Jansen [MIM:156400], resulting in impaired chondrocyte proliferation and therefore a widened growth plate.[87] On the other hand, inactivating mutations cause autosomal recessive Blomstrand chondrodysplasia (BOCD [MIM:215045]), a lethal condition characterized by severely advanced skeletal maturation, increased bone density, and severe midface hypoplasia.[88] Heterozygous mutations in *IHH* and *PTHLH* lead to short stature and brachydactyly,[89,90] while homozygous or compound heterozygous mutations in *IHH* have been reported to cause a more severe phenotype of acrocapitofemoral dysplasia (ACFD [MIM:607778]).[91]

During the maturation of bone, the hypertrophic chondrocytes secrete extracellular matrix, mostly composed of collagens and minerals, a process regulated by RUNX2, important for the formation of the periosteum and osteoblast differentiation.[92] RUNX2 regulates type X collagen expression, responsible for the deposition of extracellular matrix. Mutations in *COL10A1* are associated with autosomal dominant Schmid type metaphyseal chondrodysplasia (MCDS [MIM:156500]) and lead to altered chondrocyte differentiation.[93]

The fibroblast growth factors (FGFs) regulate many functions of the skeleton, including embryonic development, cell growth, and migration and are very important during epiphyseal growth plate formation. The complexity and diverse functions of FGFs is exemplified by the wide range of congenital disorders observed when the pathway is disturbed. A member of this family is *FGFR3,* where activating mutations in the gene give rise to a spectrum of phenotypes from mild to severe conditions, including hypochondroplasia (HCH [MIM:146000]), achondroplasia (ACH [MIM:100800]), severe achondroplasia with development delay and acanthosis nigricans (SADDAN [MIM:616482]), and thanatophoric dysplasia (TD1 [MIM:187600]; TD2 [MIM 187601]). The severity of the phenotype correlates with the degree of the receptor activation. A strong FGFR3 activation inhibits chondrocyte proliferation, hypertrophy, and increases mineralization to a more severe degree in TD1, TD2, and SADDAN, while moderate FGFR3 activation leads to milder abnormalities in the growth plate in ACH and HCH.[94] In contrast, inactivating mutations in *FGFR3* cause tall stature with arachnodactyly.[95]

1.3.1.2 Intramembranous ossification

In the intramembranous ossification the flat bones (the skull, clavicle, and lower jaw) are formed without chondrocytes as precursors. Ossification is initiated by mesenchymal stem cells (cranial neural crest cells) that form a scaffold where the cells differentiate directly into bone-forming cells, the osteoblasts. The osteoblasts produce type I collagen and osteoid matrix and mineralize. The process is initiated by the transcription factors RUNX2 and its downstream target SP7 (Osterix), both proteins important in both intramembranous and endochondral ossification.[96,97]

Heterozygous mutations in the Runt-related transcription factor 2 (*RUNX2*), which is required for osteoblast and chondrocyte differentiation, results in cleidocranial dysplasia

(CCD [MIM:119600]), a skeletal dysplasia with insufficient membranous ossification of the skull and clavicles.[92,98] Complete loss of *RUNX2* leads to poor calcification of bone in mice,[99] but in humans homozygous LOF variants in *RUNX2* have not been described. Other transcriptional factors important for appositional growth of the flat bones are *MSX2* and *TWIST*, and mutations in both genes are known to cause craniosynostosis.[100,101]

1.3.2 Bone remodelling

Bone remodelling is the process by which the old bone is resorbed by osteoclasts (bone resorption) and new bone is produced by osteoblasts (bone formation). This activity is rigorously regulated to retain a healthy skeleton and an imbalance between these two features lead to skeletal dysplasias characterized by increased or decreased bone mineral density (BMD). Impaired bone resorption is dependent on the activity of the osteoclasts and its secretion of enzymes important for resorption. It is highly regulated by calcium levels, low levels of calcium in blood prompt secretion of parathyroid hormone (PTH), which increases the activity of the osteoclasts and release of calcium from skeletal tissues. In opposite, high levels of calcium stimulate secretion of calcitonin which inhibits the activity of osteoclasts and leads to decreased bone resorption. A large number of cytokines and growth factors, such as oestrogen, growth hormones, and glucocorticoids are known to regulate bone remodeling.[102] Acquired abnormalities of the above mentioned hormones influence skeletal metabolism, often leading to secondarily decreased BMD. Primary congenital disorders with symptoms of impaired bone homeostasis are for example *osteogenesis imperfecta* (OI) and different types of osteopetroses.

OI is characterized by brittle bones and is inherited in a recessive or dominant manner and occurs when type I collagen genes or genes responsible for synthesis of type I collagen or its posttranslational modification are mutated. OI patients' bones are characterized by low BMD that easily break, leading to multiple fractures. Glycine mutations in *COL1A1* and *COL1A2* are the most common reason for OI, and these mutations act dominant negatively, disrupting the triple helix, and folding of the collagen.[55] While the bones are less dense in OI, osteopetroses have an increased BMD due to defective bone resorption, but in both diseases the bone is fragile and breaks easily. The molecular causes for osteopetroses are either mutations in genes essential for osteoclast differentiation in the RANKL signalling pathway or in genes essential for the acidification of the extracellular compartment, which disrupts the osteoclast function and leads to impaired bone resorption.[103] One example being *CLCN7,* that encodes a membrane protein responsible for the increase of chloride concentration within the lysosomes, thus an important player for the acidification.

In conclusion, there are many proteins and RNAs involved in the complex regulatory pathways important for skeletal development and homeostasis. If any of these processes are affected by a genetic abnormality, they can result in a congenital skeletal condition.

2 RESEARCH APPROACH

2.1 AIMS

The overall aim of this thesis was to expand the genetic and phenotypic spectrum of skeletal dysplasias in order understand the molecular background of some selected clinical entities and to improve diagnostic rate for skeletal dysplasia patients using WES and WGS. The specific aims addressed in this thesis were as follows:

I. To summarize clinical, radiographic, and genetic features of previously recognized skeletal dysplasias in order to obtain data on natural course and phenotypic spectrum of the conditions.
II. To identify novel clinical and genetic entities among the skeletal dysplasias and to understand their associated molecular pathology.

2.2 PATIENTS

In **study I-V** patients were referred to Department of Clinical Genetics, Karolinska University laboratory, Karolinska University Hospital because of a suspected congenital skeletal dysplasia. Informed consent was obtained from all affected patients or their parents/legal guardians and the studies were approved by the ethical review board at Karolinska Institutet. Thorough medical history, including skeletal features and symptoms, orthopaedic surgeries, extra-skeletal manifestations, medications, and anthropometry data were collected from hospital records. Skeletal surveys were reviewed at least by two experts within clinical skeletal dysplasia diagnostics. In **study III-V** patients were also referred to us from collaborating groups and included in the studies because they presented with the same clinical features. In all studies, clinical assessment and interviews of the affected patients and/or their families were performed.

2.3 METHODS

This section describes the methods used in the studies to generate data for interpretation of the congenital skeletal dysplasias.

In the past, loci for monogenic disorders were found with linkage markers and many of the mutated genes behind skeletal dysplasias were mapped with this or similar techniques.[18,20] The main techniques used in the clinic today are Sanger sequencing, multiplex ligation-dependent probe amplification (MLPA), and array-comparative genomic hybridization (aCGH). Sanger sequencing depends on PCR amplification of regions of interest, and in the end will resolve the order of nucleotides in the DNA sample to be compared to a reference sequence. MLPA and aCGH detect CNVs with a resolution depending on where the probes are located, but generally smaller deletions and duplications are not found.

Currently, exome and genome sequencing have successfully been implemented in the diagnostics of rare heterogeneous diseases, including skeletal dysplasias. The real power of this technology is forthcoming, but using whole genome sequencing data is of great

importance to find CNVs, and structural abnormalities such as inversions and translocations, in addition to SNVs and smaller deletions and duplications, that otherwise would be missed.

2.3.1 General outline of the studies

Individuals without known genetic cause for their skeletal conditions were included in subsequent studies aiming to identify causative gene defects. The collected DNA samples were analysed with WES or WGS and in selected cases with aCGH to detect or confirm gene dose imbalances. Disease-causing and candidate variants were verified using Sanger sequencing or MLPA. The applied methodology was decided on a case-by-case basis, a typical flow chart of the study can be seen in **Figure 7**.

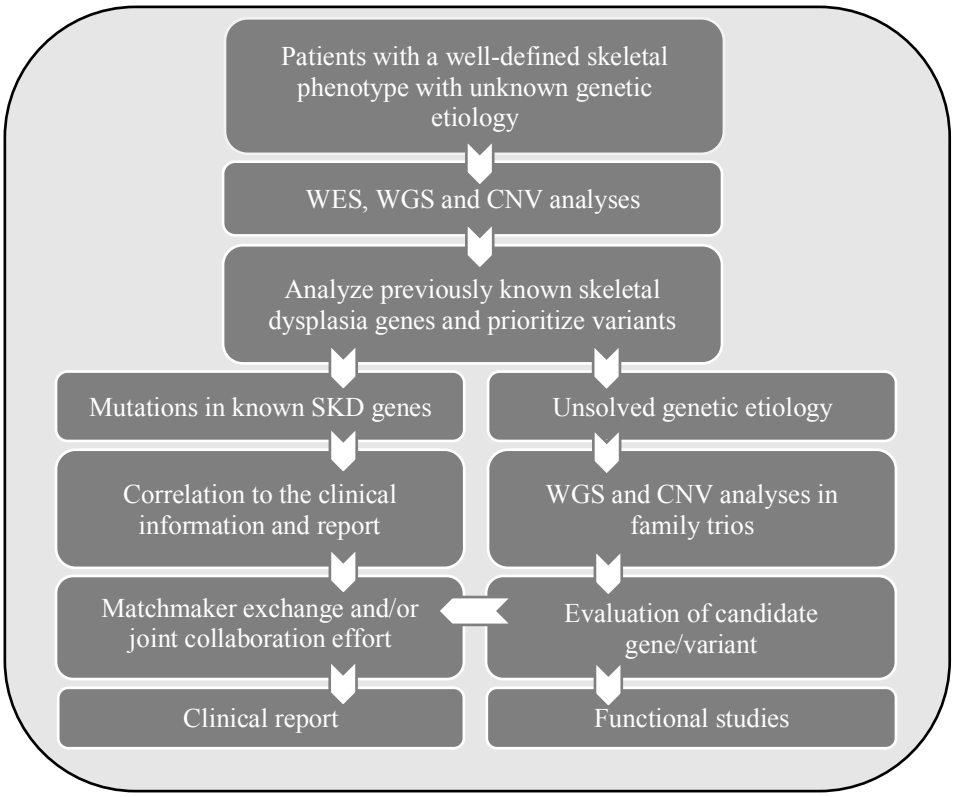

Figure 7. Flow chart of study - from patient with a skeletal dysplasia to molecular diagnosis. Many times, a joint international collaboration is needed to find more than one patient with the same rare disorder and functional studies have to be performed to confirm pathogenicity of the detected novel variants. CNV, copy number variation; SKD, skeletal dysplasia; WGS, whole genome sequencing.

2.3.2 Whole genome and exome sequencing

In brief, WES and WGS allow massive parallel sequencing simultaneously.[104-106] First, the DNA is fragmentated randomly into short segments (around 300-500 base pairs and organized into libraries. Second, adaptor sequences are ligated to both ends of the fragments and the adaptors bind to an array that allows the single-stranded DNA to be physically separated. In some applications each DNA segment is then amplified by PCR, or directly undergoes cyclic array sequencing – where primers bind to the adaptor sequences and sequencing starts. Each nucleotide carries a fluorescent label and is read directly when bound to the DNA. Both ends of the DNA are sequenced giving "paired end" (PE) reads where the number of bases between the ends is known (both reads map within a short distance). In

contrast, the mate pair sequencing (MP) has longer inserts, 2-20 kbp with PE-reads, but since a pair of inserts have a larger distance they are more suitable to read through repetitive sequences and sequences with structural rearrangements. The consensus in the field is to have a mean coverage at 30-40 times [30-40X], which reduces the error rate and reveal heterozygous variants. 30X mean coverage will cover 99.96% of the reference genome, but only 95% will be covered more than 8X.[106]

The sequences are subsequently aligned to the human reference genome with overlapping reads, followed by variant calling. Since WGS identifies within 3.4 to 5.3 million variants in an individual's genome compared to the reference, bioinformatic tools are essential. To filter the variants based on phenotype, segregation analysis, and/or online databases reduce the number of candidate variants. Another way to reduce the target region could be to only sequence the coding parts – exome sequencing – covering around 1.5% of the genome. This yields roughly 25 000 genetic variants[107] and reduces the amount of data and time needed to analyse the data. While WES is suited to identify monogenic disorders, it cannot be used to find breakpoints for structural rearrangements and the first exon usually has a lower coverage due to its high GC content, and does not target non-coding regions. WGS provides a more homogeneous coverage of the whole genome data, is more efficient and can be used to find not only single or few nucleotide variants, but also structural and gene dose abnormalities.[108] **Figure 8** shows one example of how WGS data finds a 5 kb deletion of *FOXC1* in a patient just by looking at the coverage.

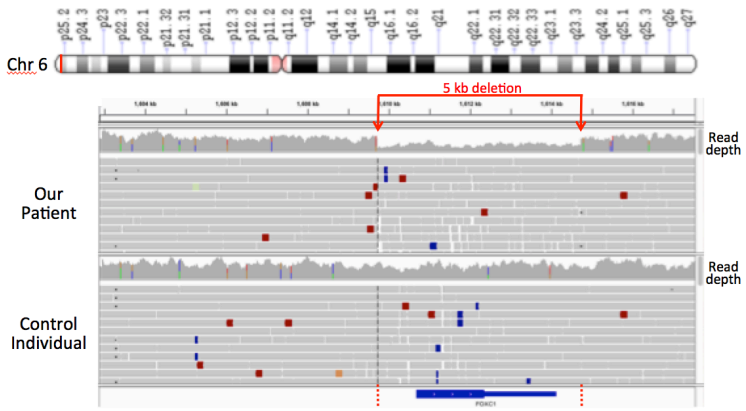

Figure 8. WGS data in proximity to *FOXC1*, showing the deletion breakpoints on chromosome 6p of a patient with a skeletal syndrome. There is only loss of one copy of *FOXC1* and no other genes are involved. This deletion was confirmed by aCGH. [Courtesy of Grigelioniene G and Garza-Flores A, unpublished data]

The interpretation of genomic variants is challenging, but the workflow to confirm the causality of the variants is:

I. Segregation analysis in the family.
II. Identification of the same disease-causing variant or other pathogenic variants in the same gene, in unrelated individuals with the same phenotype (using literature reports, DECIPHER, HGMD, ClinVar, and/or Matchmaker exchange).
III. Information of frequency of variants in our in-house database and online databases, such as gnomAD, ExAC, 1000G, and SweGen.

IV. Prediction tools of the conservation and functional effect of the variants, such as CADD, GERP, SIFT, Polyphen-2, ESEfinder, and different splice prediction tools.

The segregation analysis of a variant is usually done with Sanger sequencing of parental samples and in some cases also of other family members. Segregation analysis is used to examine if variants are following with the inheritance pattern of the studied condition in the family or if they are *de novo*. When analysing X-linked recessive disorders, it is important to discern whether the mother is a heterozygous carrier and if there are any other female members in the family that might be at risk of carrying the disease. For a VUS on the X chromosome it is useful to see if the variant can be found in hemizygous state in a healthy male relative or not. For novel variants in genes previously not known to be involved in disease, functional studies are necessary to confirm the pathogenicity.

2.3.3 Copy number variant analysis

MLPA and aCGH are normally used when a gene dose imbalance of a gene or region is suspected. With MLPA, a multiplex PCR is used that allow amplification of multiple targets in one reaction. Separating the fragments by size in each individual gives a ratio between patient and reference (healthy individual) indicating a deletion (<0.75) or a duplication (>1.25). Specific modifications of MLPA can also be used for analysis of DNA methylation and single nucleotide variations.[109] For aCGH, gene dose imbalances are found with differentially labelled DNA of a patient and a reference sample that are hybridized to an array with probes. A dosage change is found when the ratio of patient and reference fluorescent signals are unequal. It is possible to run custom designed arrays, which means that areas are selected on the array slide to cover regions associated with disease, an approach which is combined with probes distributed evenly over the genome (used in this study). Another approach for aCGH is SNP arrays that will find loss of heterozygosity (LOH) in addition to gene dose imbalances.[110]

To confirm the exact breakpoints of a deletion, a break-point PCR can be designed to amplify the region of interest, with amplification only possible when the break-point is present in the DNA sample. A break-point PCR can also be used for inversions and translocations, but since they are not CNVs, they will not be found with MLPA or aCGH, but are possible to detect with WGS.

2.3.4 cDNA analysis of predicted splice variants

The molecular diagnoses were confirmed in a total of six families (in **study I** and **V**) by analysing cDNA from the affected patients or their parents. The variants were suspected to affect splicing, either in close proximity (+2 base pairs from the exon-intron boundary) (**study I**) or further away (9 and 16 bases upstream of canonical splice site) (**study V**). Furthermore, two synonymous variants in *KIAA0753* and *DYNC2H1*, respectively, were thought to introduce exon skipping (**study V**). Primers were designed for the coding region close to the variants in *ALG9*, *DYNC2H1*, *EVC*, and *KIAA0753* (one of the primers usually spanning over two exons to reduce amplified genomic DNA) and subsequent Sanger sequencing was performed to show the aberrant splicing.

2.3.5 Histology and immunohistochemistry of the growth plate

Growth plate studies are usually only possible in animal models and in fetuses from terminated pregnancies. These samples can be used to observe structural abnormalities of the growth plate and to examine expression of RNA or proteins of interest. In **study IV** paraffin-embedded sections of the distal femoral growth plate from the affected fetus was stained with haematoxylin-eosin and compared to an age-matched control fetus. Further, immunohistochemistry with antigen retrieval was performed in citrate buffer and after blocking the slides they were incubated with primary antibody to KIAA0753 and GAPDH, secondary fluorescent antibody and nucleus was counterstained with DAPI. Images from the microscopy were analysed using ImageJ software.

2.3.6 Zebrafish as a disease model

In order to verify the disease-causing effect and to study the pathophysiology of a specific skeletal dysplasia gene, we can use model organisms. This is possible since many genes are conserved between species and their proteins have similar functions. Zebrafish are particularly suitable for early developmental studies, because of its external fertilisation and development (compared to *in utero* in mice). We can therefore, follow the embryonic development using a dissection microscope. Other advantages of using zebrafish in genetic studies include the large collection of mutant and transgenic lines and the molecular tools available to knock-down or overexpress genes of interest. Tissues and organs from zebrafish and humans are in general comparable, especially at the cellular level, and therefore zebrafish can be used to understand specific pathways and mechanisms, including cilia biogenesis. Although, some molecular mechanisms may be different between them similar phenotypes have been observed in zebrafish when knocking out disease genes and give a model that resembles ciliopathies in humans.[111] By comparing wildtype and mutant zebrafish, we can describe the role of any gene of interest, its role in disease, and test compounds with therapeutic potential.

In **study IV** zebrafish was used as a model organism to confirm the role of *KIAA0753* in a skeletal ciliopathy syndrome. The zebrafish genome shares more than 70% of their genes with humans and is therefore a good system to study many human diseases.[112] For this study, the zebrafish line *sa22657,* which contains a premature nonsense mutation in *Kiaa0753*, was used. This line was generated within the zebrafish mutation project (ZMP),[113] established by the Wellcome Trust Sanger Institute, and ordered from the European zebrafish resource center (EZRC).[114] ZMP was established to generate knockout alleles of all genes in the zebrafish genome. Mutagenesis was performed by exposure of male zebrafish to ENU, a mutagenesis agent and progeny screened for mutations. Single-mutation lines have been established by outcrossing mutant specimens with wildtype zebrafish and genotyping. The *sa22657* zebrafish line was maintained at the Karolinska zebrafish core facility. Zebrafish morphology examination was performed using a dissection microscope on five days post fertilisation zebrafish larvae, by staining with Alcian Blue.[115]

2.3.7 Biochemical tests for CDG

Skeletal symptoms can occur in different metabolic conditions, such as abnormal GAG metabolism in dysostosis multiplex, insufficient protein glycosylation in CDG or elevated levels of cholesterol precursors in X-linked chondrodysplasia punctata.

In **study I** we found potential disease-causing variants in *ALG9* and evaluated the abnormal glycosylation pattern of transferrin in the patients and their parents. Transferrin is a highly abundant serum glycoprotein synthesized in the liver and is important for iron transport. It can be used as a biochemical marker for CDG, since they have elevated concentrations of carbohydrate-deficient transferrin (CDT). Glycosylation tests can be performed using CDT analysis to determine the defect glycosylation pattern of transferrin, using isoelectric focusing (IEF), capillary zone electrophoresis (CZE) or mass spectrometry. In this study we used liquid mass spectrometry analysis.

2.3.7.1 Liquid chromatography-mass spectrometry analysis of transferrin

CDGs are a group of disorders with defective enzymes in the biosynthesis of glycoproteins. The asparagine-linked (N-linked) CDGs can be tested in clinical practice by looking at the abnormal pattern of N-linked oligosaccharides of serum transferrin, which is the most sensitive marker of a defective N-glycosylation.[38] In **study I**, serum from the affected fetuses was not available, therefore biopsies from frozen spleen were used instead.

Liquid chromatography-mass spectrometry (LC-MS) method was used to demonstrate the abnormal isoforms of transferrin.[116] In brief, aliquots of the samples were applied to an immunoaffinity column and enriched transferrin was eluted and concentrated on an analytical C_4 column. Transferrin was eluted from the C_4 column and preceded with mass spectrometry. Analysis of transferrin was done using frozen spleen tissue from patients and from age-matched controls and one adult. The calculation of the ratios of mono-glycosylated/di-glycosylated transferrin can then be compared between the samples. The ratio was significantly elevated in the patients compared with age-matched controls, and showed that the affected individuals had underglycosylated transferrin.

LC-MS is a sensitive and automated method to determine the defect N-glycosylation in CDG. Abnormal hypoglycosylation patterns can also be seen in acquired hypoglycosylation (secondary), such as galactosemia, hereditary fructose intolerance or moderate to heavy alcohol consumption.[117] CDG analysis due to a congenital disease is typically performed at young age so there is little risk of false positives, but the profile is the same for acquired hypoglycosylation and relevant clinical information may be necessary to interpret the results correctly.

3 RESULTS & DISCUSSION

3.1 EXPANDING THE GENETIC SPECTRUM OF SKELETAL DYSPLASIAS

It is not always obvious that possible pathogenic variants detected by sequencing have a role in bone biology and often the molecular mechanisms of conditions are not completely understood at the time of first report. For instance, disease-causing variants in *ALG9* in **study I** and *KIAA0753* in **study IV** were not previously known to be associated with skeletal abnormalities and therefore the role of these genes in bone biology were not clear. In such cases, several unrelated patients with the same phenotype and disease-causing variants in the same gene are needed to establish the proofs. Furthermore, functional validation of those variants and tissue expression of candidate genes are necessary, using cells cultures, tissues or model organisms. In **study II** *COL2A1*, **study III** *BMPER*, and **study V** *C2CD3, DYNC2H1, EVC, EVC2*, and *WDR19* were genes already known to be involved in congenital skeletal disorders and bone biology, therefore easier to interpret variants in those genes in our patients.

In some consanguineous families, such as family 2 in **study I** and family 1 in **study IV**, patients may have multiple pathogenic variants in known disease-associated genes. In **study I**, we found that affected fetuses from family 2 in addition to the pathogenic variants in *ALG9,* were also homozygous for potentially disease-causing variants in *ANK3*. Pathogenic variants in the *ANK3* gene have been associated with autosomal recessive mental retardation 37 (MRT37 [MIM:615493]).[118] We reported the variant in *ANK3* as a VUS, which most likely does not contribute to the phenotype of *ALG9*-CDG in this family.

Further, two affected siblings from family 1 in **study IV** were homozygous for pathogenic variants in *KIAA0753* and *SLC13A5*, respectively. The findings of two rare conditions rather than one disease with phenotypic expansion affecting these individuals, is not an unusual situation in consanguineous families or isolated populations.[78-80] On the other hand, the variant in *SLC13A5* in family 1 (**study IV)** had previously been reported to cause autosomal recessive early infantile epileptic encephalopathy 25 (EIEE25 [MIM:615905]), a condition associated with seizures, developmental disability, and teeth hypoplasia.[119,120] These features were observed in the affected siblings, but since two unrelated families had the same homozygous nonsense variant in *KIAA0753* and did not share the *SLC13A5* variants, we concluded that the seizures and teeth hypoplasia were due to the homozygous *SLC13A5* variant. The developmental and speech delay are overlapping features and could be inferable to either one of the congenital conditions. In conclusion, in some cases with new features reported in association to a certain condition, there might not always be a phenotypic expansion of a disease, but two genetically independent conditions. Extended SNP analysis indicated that the two families did not have a common ancestor.

3.2 NOVEL DISEASE GENES CAUSING SKELETAL DYSPLASIAS

There are numerous skeletal dysplasias with unknown genetic causes. One example is a very rare lethal fetal syndrome with polycystic kidneys, typical facial features and varying malformations with severe skeletal dysplasia were first characterized by Gillessen-Kaesbach *et al.* in 1993 and Nishimura *et al* in 1998.[121,122] In **study I**, we identified three fetuses from two unrelated families affected with the same condition and show that this syndrome belongs to the most severe spectrum of N-glycosylation disorders, caused by homozygous protein truncating variants in *ALG9*.

The gene, *ALG9* encodes for the alpha-1,2-mannosyltransferase, an enzyme involved in the formation of the lipid-linked oligosaccharide (LLO) precursor of N-glycosylation. Previously, reported patients with ALG9-CDG had missense variants and suffered from intellectual disability, muscular hypotonia, microcephaly, and small renal cysts, but skeletal dysplasia features were not described. However, skeletal abnormalities, similar to our patients with ALG9-CDG have been described in ALG3- and ALG12-CDG,[123-125] suggesting that this subset of glycosylation disorders constitutes a new diagnostic subgroup of skeletal dysplasias.

This group of enzymes (ALG3, ALG9, and ALG12) are all important in adding mannose residues to the LLO in four direct sequential steps in the biosynthesis of LLOs, in the luminal side of the ER (**Figure 2**). Our study shows that protein truncating variants in *ALG9* can lead to a lethal skeletal dysplasia and visceral malformations as the most severe phenotype associated with defect N-glycosylation. Another fetus, born to unrelated parents and not included in our study, but clinically diagnosed with similar skeletal features and previously reported by Kranz *et al.*,[125] turned out to have two protein-truncating variants in *ALG3* (p.Met1? and p.Arg117*) (Hammarsjö and Zerres unpublished data).

The most severe phenotype of ALG9-CDG is now included in the OMIM database as Gillessen-Kaesbach-Nishimura syndrome (GIKANIS [MIM:263210]) and is an important differential diagnosis to many fetal lethal malformation syndromes. Since our report, another three families (6 patients) with ALG9-CDG with abnormal N-glycosylation were reported, all with homozygous missense mutations and no features of severe skeletal dysplasia.[126-128] Nonetheless, even though there are only sixteen patients reported so far with ALG9-CDG the splice variant in our three patients is associated with a severe skeletal phenotype.

Skeletal ciliopathies are a rare group of diseases, divided into different subtypes with a significant genetic heterogeneity and phenotypic overlap.[28] **Study IV** revealed four patients and **study V** reported two siblings with skeletal ciliopathy phenotype and protein truncating variants in *KIAA0753*. Pathogenic variants in this gene were previously described to cause ciliopathies like Joubert (JBTS)[129] and orofaciodigital syndrome (OFD),[130] but no skeletal symptoms were reported in those three individuals. It is, therefore, not surprising that four of our patients (family 1 in **study IV** and family 13 in **study V**) showed brain abnormalities consistent with JBTS and fetus in **study IV** had features resembling OFD.

Examination of the femoral growth plate from the affected fetus showed an abnormal architecture of the proliferative zone with lack of normal chondrocyte columns, consistent

with skeletal ciliopathy syndromes.[31,131] Immunohistochemistry of the growth plate from normal age-matched control and affected fetus showed that KIAA0753 is expressed in the growth plate from normal fetus, but absent in the corresponding cells from the affected fetus.

A zebrafish knockout model, carrying a homozygous nonsense variant in the *kiaa0753* gene, showed classic ciliopathy phenotype with cartilage abnormalities, such as curved body and an abnormal head morphology due to altered cartilage patterning. The morphological assessment could only be performed at early developmental stages, because of early lethality of the larvae, not surviving beyond the first week of development. Overall, the structure of the skull was compressed along the anterior-posterior axis when compared to wildtype siblings. These results supported the involvement of KIAA0753 in the skeletal morphogenesis.

In the small cohort of patients with *KIAA0753*-related syndrome that have been reported so far, our patients seem to have symptoms at the most severe end of the phenotypic spectrum. We hypothesize that the skeletal abnormalities in our patients are due to PTVs, leading to absent KIAA0753 expression and that the milder phenotypes of the previously reported patients may be due to some residual expression or milder effect of KIAA0753 function.

3.3 DETECTING NEW VARIANTS IN KNOWN DISEASE GENES

In **study II**, we identified a missense variant in *COL2A1* (c.3655G>C, p.Asp1219His) in six affected family members with spondyloepiphyseal dysplasia type Stanescu (SEDSTN [MIM:616583]), and early onset of pain and deformities in the joints. Previous publications have reported similarities between SEDSTN and Czech dysplasia [MIM:609162], which is caused by a recurrent mutation in *COL2A1* (c.823C>T, p.Arg275His), but distinguished from SEDSTN by the hallmarks of short 4th and 5th metatarsals.[132] **Study II** extends the phenotypic spectrum of collagenopathies type II. Simultaneously another study reported three individuals from two unrelated families with SEDSTN and a missense variant in *COL2A1* (c.619G>A, p.Gly207Arg)[133]. Correlation of genotype-phenotype in SEDSTN is not possible so far, because only a few patients have been reported to date. It is, however, relatively well established that there is no good genotype-phenotype correlation in *COL2A1*-related skeletal dysplasias.[134,135]

Most mutations in *COL2A1* are in the triple helix domain and a majority affect a glycine residue (every third amino acid in the alpha helix is a glycine, Gly-X-Y), required for a correct conformation of the collagen fibre and the glycine mutations are dominant-negative. The type II collagenopathies are mainly inherited in an autosomal dominant matter, but in two studies (**related publication VIII** and Barat-Houari *et al.*) have found a severe spondyloepiphyseal dysplasia congenita (SEDC [MIM 183900]) with homozygous *COL2A1* non-glycine missense mutations.[136] Typically, non-glycine missense mutations, as well as truncating variants, are usually associated with milder phenotype. However, in these two families the missense mutation in heterozygosity did not result in any clinical symptoms.

In **study III** we identified biallelic variants in *BMPER* in two patients with ischiospinal dysostosis (ISD), a disorder with abnormal ossification of the spine. The causative gene had

so far been unknown, but the condition was suggested to be of autosomal recessive inheritance and allelic to the lethal condition diaphanospondylodysostosis [MIM:608022].[137] Diaphanospondylodysostosis is caused by pathogenic variants in the *BMPER* gene, the symptoms are similar to, but more severe than those of ISD.[137] Our study confirmed the hypothesis that ISD and diaphanospondylodysostosis were allelic disorders and extending the phenotypic variabilities in *BMPER*-related skeletal dysplasias.

For patient 1, WES and segregation analyses revealed compound heterozygous mutations in *BMPER,* a highly conserved missense variant and a nonsense mutation. Patient 2 had three variants in *BMPER*, a homozygous nonsense and predicted deleterious missense variant (the mother carried both the nonsense and the missense variant). The missense variant was predicted not to be involved in this patient's phenotype, because the nonsense variant was located upstream. The missense variant in patient 1 could explain her milder phenotype, but in background of the previously reported diaphanospondylodysostosis variants, the genotype-phenotype correlation is not clear for BMPER-related disorders. Subsequently, further reports indicated that diaphanospondylodysostosis and ISD should be considered as a phenotypic continuum. Legare *et al*. described a 9-year-old boy first classified as ISD, but later reclassified to diaphanospondylodysostosis due to his severe phenotype. This patient was compound heterozygous for a deletion that included *BMPER* (and eight other genes) and a missense variant.[138] The fact that there was no renal dysplasia or Wilms tumour in the two patients in **study III** or in the patient reported by Legare *et al*. might contribute to their long-term survival. Having in mind that diaphanospondylodysostosis is associated with Wilms tumour, individuals with ISD should be considered regular follow-up with ultrasound during childhood in case milder mutations lead to increased tumour risk later on.

Another group of dysplasias we studied are skeletal ciliopathies which show a broad phenotype and genotype variability. Deleterious variants in genes coding for different ciliary components affect the function of the cilia in a similar way, leading to overlapping phenotypes and *vice versa*, with mutations in the same gene may lead to a broad phenotypic spectrum.[1,28] In **study V**, we investigated the genetic causes of skeletal ciliopathies in 24 unrelated individuals and identified disease-causing variants in 83% of the cohort.

Strikingly, four of the patients had splice events outside the exon-intron boundaries (±2 nucleotides) and RNA from them was needed to prove the possible pathogenicity of the variants. Two of the variants were seemingly synonymous and would have been discarded as benign if it wasn't the fact that they segregated in the families and had an extremely low MAF (<0.0001). None of the interpretation software used could predict that these variants disrupted an exonic splicing enhancer or silencer (ESE or ESS). However, Sanger sequencing of cDNA showed that both synonymous variants in *DYNC2H1* and *KIAA0753*, respectively, led to aberrant splicing due to exon skipping and frameshift predicted to result in NMD.[77]

Two of the variants in close proximity (9 and 16 nucleotides upstream) to the canonical splice sites were shown to introduce cryptic splice sites (novel AG dinucleotides), in *EVC* and *DYNC2H1*. These rare intronic variants were both indicated by splice prediction tools to introduce new acceptor splice sites, leading to retention of intronic sequence within the

transcript and frameshift with predicted NMD. We pursued with Sanger sequencing of cDNA from fetal lung (*DYNC2H1*) and blood samples from heterozygous mother (*EVC*).

All of the four patients' splice variants were found when re-evaluating the sequence data, showing a need for regular reanalyses of previously unsolved cases. Although the distant splice variants or pathogenic synonymous variants are hard to interpret, the workflow changes with new *in silico* prediction tools and reporting disease-causing variants in databases (like HGMD, ClinVar, Matchmaker Exchange, DECIPHER), makes it possible to redeem genes and variants of unknown significance.

Structural analysis showed that one patient with phenotype of mild skeletal ciliopathy had a 1q24q25 microdeletion syndrome and another with SRTD harboured an intragenic deletion within *DYNC2H1*, which displays the importance of performing gene dose analysis and to include possible differential diagnoses.

In four of the patients in **study V** no molecular diagnosis could be determined, although extensive analyses were performed with WGS trio sequencing. This may be attributed to the fact that disease-causing variants in intronic and regulatory regions are challenging to interpret using the current methodology or that there are novel genes not yet associated with the clinical conditions. Future studies, integrating RNA-sequencing may be helpful in identifying pathogenic variants in these patients, if RNA is made available. Here we were able to show that it is possible to find splice variants in RNA from blood from individuals with ciliopathies and it may be possible also for the unsolved cases. There have been a number of publications showing that there is a great increase in the diagnostic yield using RNA-sequencing to find variants causing abnormal splicing.[139-141] Clinical correlation of molecular findings and clinical phenotyping is necessary to accurately interpret variants of unknown clinical significance, as well as to improve accuracy of bioinformatic tools.

3.4 SKELETAL DYSPLASIA GENE PANEL IN A CLINICAL SETTING

In a clinical setting we have established an *in silico* gene panel, to date, comprising of 466 genes which are known to cause congenital skeletal disorders and have analysed DNA samples from 178 skeletal dysplasia patients (including 21 probands from **study II, III**, and **V**) (unpublished data). The analyses are performed using a visualization interface of genomic data, Scout software,[142] developed by Clinical Genomics at SciLifeLab in collaboration with the clinical departments at Karolinska University Laboratory.

In this cohort, spanning many of the groups from the skeletal nosology,[1] 51% of the patients received genetic diagnoses using WGS or WES data. The patients in the study include children (n=115), adults (n=33), and fetuses (n=30). There is no difference in genetic diagnosis outcome between the adult patients or children, but diagnostic yield is higher in the fetal cases (67%), which may reflect the severity of disease.

Variants were found in 50 established skeletal dysplasia genes and all have been confirmed with Sanger sequencing and segregation analysis performed if parental samples were available. While there were many variants that could be classified as disease-causing, our

analysis revealed 10% of the variants of unknown significance in known disease genes that might explain the skeletal abnormalities of the patients. Further functional studies or extended RNA analyses are needed to confirm or discard pathogenicity. The analysis is mainly done as a singleton with only the affected individual, however in some cases sequencing of multiple samples from several unrelated families have been added to identify the disease-causing variants. The remaining patients without a genetic diagnosis might need family trio sequencing or RNA sequencing to reach the genetic diagnosis.

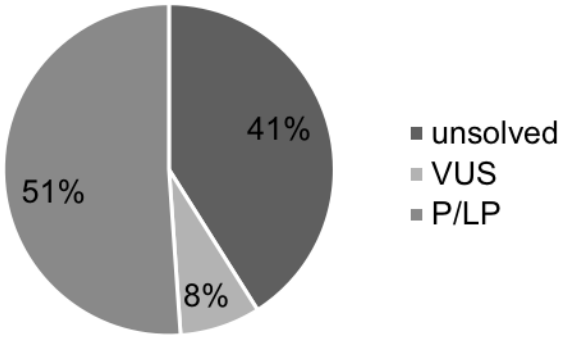

Figure 9: The frequencies of genetic diagnoses 178 skeletal dysplasia patients using an *in silico* gene panel from WES or WGS data. The 178 patients were referred to the clinical genetics department at Karolinska University Hospital during 2014 to spring 2018. 91 patients had a disease-causing variant in a skeletal dysplasia gene, 14 patients had a VUS (that are under current investigation and has not been reported to the patients) and 73 patients have not yet received a genetic diagnosis. LP, likely pathogenic; P, pathogenic; VUS, variant of unknown significance; WES, whole exome sequencing; WGS, whole genome sequencing.

We should highlight that RNA analyses of candidate splice variants (outside the canonical splice site) are of great importance to be able to confirm them as disease-causing (8% of the positive findings). Another important aspect is that nine patients (10%) had structural abnormalities that explained the phenotype. In the future, this means that WGS data is of great importance to detect structural variants.

For the consanguineous families, regions of homozygosity can help to elucidate pathogenic variants that are shared between family members. In this cohort 25 of the cases were from consanguineous families and 84% got a genetic diagnosis, all but three were homozygous for a pathogenic variant. These individuals were found to have known disease-causing variants in *FGFR3*, *SATB2*, and *FLNA*, indicates that even though consanguineous families have a higher incidence of autosomal recessive disorders, *de novo* variants or X-linked recessive variants still occur and should be considered.

In the future, when costs decline and new pipelines emerge the WGS is likely to be the first-tier analysis method, for both single gene disorders and structural variants. This will increase the diagnostic yield and also make it possible the re-evaluate the data from individuals with unsolved diagnoses and add family members for trio analysis.

3.5 LIMITATIONS

A major limitation in this type of studies is that functional studies are often needed in order to examine the consequences of very rare variants, and to explore the underlying mechanisms and pathophysiological pathways of the conditions. In approximately half of the patients with skeletal dysplasias we fail to achieve the genetic diagnosis, despite using WES or WGS. This could be due to novel disease genes or variants in non-coding regions, splice aberrations, structural variants, repeat expansions, and epigenetic modifications.

For patients where singleton WGS does not provide a genetic diagnosis, adding family members to the sequencing analysis may be helpful. If a novel gene is identified, increased sample size of patients is needed to enable the new genetic diagnosis or to make genotype-phenotype correlations, but since the disorders often are extremely rare this is challenging. Matchmaker exchange and tight international collaborations makes such efforts possible (as in **study III, IV**, and **V**). Matchmaker exchange is an initiative to connect researchers and match patients with the same phenotype and genetic architecture for interpretation.[143,144] Another strategy to elucidate the genetic cause in patients with an unsolved genetic skeletal dysplasia presenting with a known phenotype is to search for possibly pathogenic variants in genes coding for molecules in the same pathways. This was done when we recently identified a novel gene behind Stuwe-Wiedemann syndrome (STWS [MIM:601559]) (Grigelioniene, Hammarsjö et al. manuscript in preparation). Another way to solve human genetic diagnoses is to recognize the malformation pattern in experimental animals and draw parallels with human conditions. Some skeletal dysplasias were genetically solved with such a hypothesis for example *TRIP11* mutations in achondrogenesis type 1A (ACG1A [MIM:200600]) were identified when it was recognized that Trip11-null mice shared skeletal similarities with ACG1A patients.[145]

How to report variants of unknown significance back to an affected family is not always simple, but ACMG states a clinical analysis of 56 genes that should be reported to the patient/family irrespective of age. Fetal and newborn screening is not included in this list.[146] The genes included are associated with conditions that are preventable or treatable when detected early enough, such as cancer susceptibility syndromes, cardiomyopathies, or aortic and intracranial aneurysms. The variants are reported back as secondary findings and should described as pathogenic or expected pathogenic based on the type of variation. Green *et. al*[146] calculated that around 1% of the clinical WES/WGS will result in incidental finding unrelated to condition of the patient.

Even if we mainly use *in silico* gene panels from WGS data, we still encounter incidental findings that are not related to the phenotype that are classified as possibly pathogenic. Often it is one variant that is associated with a recessive skeletal condition, i.e. heterozygous carrier status that we don't report to the patient. In our study of 178 skeletal dysplasia patients, we had one family with a fetus with severe OI and a *de novo* variant in *COL1A1*. Furthermore, we identified a pathogenic variant in *COL11A2* associated with autosomal dominant deafness with onset in middle age. After genetic counselling with the family and ethical considerations, we chose to report this variant to the family, since it might spare the carriers from unnecessary medical examinations at onset and maybe important when selecting occupation, and/or adjusting life style.

There are still many skeletal dysplasia patients in our clinical cohort with not yet identified genetic background. It might be due to another layer of complexity that we are not clinically ready to approach, i.e. the non-coding variants in the regulatory elements or the variants affecting the recognition of splice sites. Complex regulatory mechanisms of skeletal development, translational, and epigenetic regulation, as well as post-transcriptional modification are processes which will be highlighted by emerging new molecular techniques.

4 CONCLUDING REMARKS & FUTURE PERSPECTIVES

Complex networks are vital for the processes of bone growth and remodelling, and genetic studies of skeletal disorders have shed light on the signalling pathways involved. Since most of the skeletal dysplasias are caused by mutations in a single gene, they represent 'experiments of nature' with great potential for advancing the understanding of the specific genes involved in skeletal development. The details how the products of these genes interact in complex networks are still emerging. As indicated, skeletal radiology and pattern recognition are essential to group and recognize the conditions. Compared to facial dysmorphology, skeletal pattern recognition is an easier instrument to classify the conditions, because the radiographic morphology is less influenced by ethnical variations.

As shown for all the studies in this thesis, the genetic analyses are based on MPS technology and the advances in technology have changed the field of clinical genetics. From the beginning of 2014 the data was obtained mainly with WES, due to cost and amount of DNA needed. But as time passes the WGS costs have decreased and is now the preferred strategy (almost all-in-one-analysis). The innovation of new annotation and prediction tools that will combine and interpret the data will lead to increased number of solved genetic diagnoses. Tight collaboration between the clinicians, molecular biologists, and bioinformaticians will help to further improve the bioinformatic pipelines.

WGS is a fast and reliable method to find the causative variant for not only skeletal dysplasias, but for all Mendelian disorders. Already now the WGS data provide us with information about structural abnormalities and sequence variants with *in silico* prediction tools, evolutionary conservation, and allele frequency from large databases to interpret the variants in relation to the disease phenotype. In the near future, WGS will most likely replace the conventional techniques used in a clinical genetics laboratory today, such as karyotyping, aCGH, MLPA, and Sanger sequencing. Those methods will be used more to verify the WGS results.

Long-read sequencing is an upcoming technology that will enable us to resolve compound heterozygous variants without the need for segregation analysis, as well as finding middle-sized structural variants missed today due to technical reasons. RNA-sequencing and proteomics will also be valuable to elucidate which genomic variants actually affect splicing and will contribute in genotype-phenotype correlations. Although RNA-sequencing data is limited to the patient material and time-point it was collected, it could be used to address the problem. RNA from affected tissue is preferential to ensure that the gene involved in the pathogenesis is expressed. However, many genes are expressed in multiple tissues and we have seen that examining RNA extracted from blood is sufficient to show aberrant splicing in several cilia encoding genes. Another source to examine RNA would be from fibroblast cultures, since skeletal and cartilage tissues are difficult to obtain.

Genes important for skeletal formation are not only conserved in different species on the sequence level but also in their molecular function, which makes species like zebrafish and mice suitable to model human skeletal disorders. Discovering signalling pathways and molecules important in skeletal dysplasias can also be studied in a cellular level and the

transgenic animals or cells treated with different substances to target the affected molecular pathways. The CRISPR/Cas9 technology has been of great importance to target genes and introduce specific variants, in order to mimic disease in model organisms.

The genetic studies of monogenic diseases increase our understanding of the pathogenesis of skeletal dysplasias and the roles of these genes in normal skeletal development. New findings lead to a picture of a complicated network and reveal targets for potential new drug therapies, not only for rare skeletal disorders but also for common conditions that display the same or similar pathogenetic mechanisms, such as osteoporosis, scoliosis, and arthritis. One of the main challenges for the future are to map the genetic pathways in detail and describe the pathophysiological processes behind the skeletal diseases and find the possibilities to develop novel therapeutic strategies. There have been some convincing studies in both humans and mice to increase bone mass with monoclonal antibodies against sclerostin[147] and mice with achondroplasia have been treated with a CNP analogue (CNP antagonize the Fgfr3 signalling).[148]

Genetics are displayed in different layers, from the genomic DNA, to the more hidden information that influence inheritance, development, and disease, such as the transcriptome, the proteome, and epigenetics. Also, the non-coding RNAs are very important for the normal development to function properly. At first, the non-coding DNA was thought to make "no sense" and only the DNA coding for proteins were important, but as it appears the molecular biology is unfolding and more complex regulatory mechanisms are identified.

In the future, with WGS data from an individual we could also include an analysis to make a report of variants that could be of interest to characterize a patients' response to treatment and a list of common disease-variants, such as *CFTR* with a high allele frequency in our population, to avoid common congenital disorders in a specific population. Further, the ACMG guidelines states that secondary findings that are preventative or treatable for early detection should be brought back to the patient, regardless of autonomy because of family benefit.

This thesis has shed light on the molecular background on some of the skeletal dysplasias and given many patients their genetic diagnosis. In the future, I hope that this work add value to experimental work to develop and optimize treatment for skeletal dysplasias.

5 POPULÄRVETENSKAPLIG SAMMANFATTNING

Medfödda skelettsjukdomar innebär att skelettet har en onormal form och/eller hållfasthet, med varierande symptom som omfattar allt från fosterdöd till kortvuxenhet i vuxen ålder. Många diagnoser är associerade med andra symptom såsom neurologiska komplikationer, missbildningar och skador på inre organ. De orsakas av förändringar i arvsanlag (gener) som styr skelettets utveckling. Symptombilden varierar beroende på vilken gen som orsakar sjukdomen, dvs vilken gen som är muterad. Kliniska diagnosen kan försvåras för många skelettsjukdomar, då mutationer i en gen kan leda till flera olika sjukdomar och mutationer i flera olika gener kan leda till en och samma diagnos.

Skelettsjukdomar är i det stora hela ganska vanliga i en population och drabbar cirka en på tre tusen individer, men varje enskild diagnos är sällsynt och det saknas ofta klinisk kunskap om komplikationer och behandling. Man säger att skelettsjukdomar är en heterogen grupp då det finns en stor klinisk variabilitet där diagnosen ställs med kliniska fynd och röntgen. I möjligaste mån vill man även bekräfta diagnosen genetiskt för att på så sätt vara säker på att patienten får optimal medicinsk uppföljning.

Den nuvarande klassificeringen av medfödda skelettsjukdomar från 2015 omfattas av 436 diagnoser fördelade i 42 grupper. För många av dessa diagnoser är den genetiska orsaken ännu inte identifierad. I klassificeringen är 364 gener beskrivna, men vi känner idag till ungefär 460 gener där mutationer i dessa kan orsaka en skelettsjukdom. Tekniken för att hitta dessa mutationer bygger framför allt på massiv parallell sekvensering och är en mycket effektiv metod för att hitta den genetiska orsaken hos patienter med sällsynta genetiska sjukdomar, inte bara för sjukdomar som drabbar skelettet.

Människan har ungefär 21 000 gener, som är uppbyggda av DNA molekyler och utgör instruktionerna för att skapa liv. Det är denna genetiska kod som otroligt nog endast består av fyra kvävebaser; adenin (A), cytosin (C), guanin (G) och tymin (T) som förs vidare från generation till generation. Med storskalig sekvensering kan man bestämma ordningen på i stort sett alla tre miljarder baser i den genetiska koden (även kallad genomet). Det gör att det nu är möjligt att jämföra DNA från en sjuk individ med en referens bestående av friska individer. Användandet av storskalig sekvensering har lett till att arvsmassan hos mer än hundra tusen individer har kartlagts och användningsområdet är stort, men framför allt är den genetiska variationen tillgänglig för att avgöra vilka varianter som är mutationer och leder till en genetisk sjukdom. Medan DNA finns i cellens kärna och lagrar den genetiska koden, översätter RNA information från den genetiska koden för att skapa ett funktionellt protein.

I den här studien har vi identifierat gener som är viktiga i skelettets uppbyggnad. Det har vi gjort genom att sekvensera alla gener hos flera individer från samma familj, för att kunna skilja de sjukdomsorsakande mutationerna från normal variation (de varianter som gör oss unika). I **studie I** identifierade vi mutationer i *ALG9* genen hos två familjer som drabbats av upprepad fosterdöd, dessa var nedärvda från varsin förälder och således är en autosomalt recessiv sjukdom. Det innebär att dessa två familjer har en hög upprepningsrisk att få barn med denna svåra skelettsjukdom. Även i **studie IV** hittade vi mutationer i en gen, *KIAA0753*, som inte tidigare varit associerad med skelettavvikelser. De drabbade individerna hade en

bred variabilitet, från fosterdöd till mycket smal bröstkorg, korta ben och armar och utvecklingsförsening. Både *ALG9* och *KIAA0753* var tidigare kända att orsaka andra genetiska sjukdomar utan skelettpåverkan.

I ytterligare två studier har vi hittat mutationer i tidigare kända skelettgener, men där genetisk orsak för skelettsjukdomen varit okänd. I **studie II** har vi beskrivit symptom hos en stor svensk familj med sju drabbade individer med spondyloepifysär dysplasi typ Stanescu och visat att det beror på en mutation i genen *COL2A1*, som kodar för kollagen och är viktig i stabiliteten av brosk/ben. I **studie III** har vi två patienter drabbade av ischiospinal dysostos (ISD) med mutationer i *BMPER*. Genen är känd att orsaka en dödlig skelettdysplasi, diaphanospondylodysostos, men våra patienter är två respektive sjutton år gamla. Vi visar med denna studie att ISD och diaphanospondylodysostos är en del av samma sjukdomsspektrum och orsakas av mutationer i samma gen.

I **studie V** genomfördes storskalig sekvensering för att hitta den genetiska orsaken hos 24 patienter med skelettfenotyp karaktäriserad av korta revben och smal bröstkorg, samt korta rörben. Sjukdomen nedärvs autosomalt recessivt. Vi identifierade mutationer hos 20 av de 24 individerna (83%), där det visade sig att fyra av mutationerna behövde analyseras med hjälp av RNA för att visa att de var så kallade splitsningsmutationer, som förutses ge ett felaktigt uttryck av proteinet.

Vidare har vi undersökt närmare 200 individer med medfödda skelettsjukdomar med storskalig sekvensering av hela genomet, men där vi egentligen bara analyserar en panel av skelettgener. Vi hittar då den molekylära orsaken i ungefär hälften av fallen. De flesta individerna med genetisk diagnos har mindre sekvensavvikelser, såsom ett utbyte av kvävebas, men även gendosavvikelser (för mycket eller för lite genetiskt material) och inversioner som bryter genens funktion har hittats (se figur 3). Flertalet av de individer utan mutationer i de kända generna har unika sjukdomstillstånd, som utgör en grund för fortsatt analys för att hitta nya varianter eller gener som orsakar sjukdomen. För att analysera konsekvenserna och kartlägga mekanismer av nya genetiska varianter kan sjukdomsmodeller eller RNA studeras, såsom zebrafisk i **studie IV** och RNA-sekvensering i **studie I** och **studie V**.

Syftet med denna studie har varit att studera genetiken bakom medfödda skelettsjukdomar, där de nya genetiska orsakerna som upptäckts ökar kunskapen om de enskilda diagnoserna och till en ökad förståelse av nya sjukdomsmekanismer för reglering av skelettets tillväxt. På lång sikt kan den nya kunskapen leda till utveckling av nya läkemedel för behandling av vanliga skelettsjukdomar som benskörhet och artros. En snabb implementering av resultaten gör att en molekylär diagnos kan användas vid familjeplanering för de drabbade familjerna för att undvika upprepning av samma diagnos i familjen.

6 ACKNOWLEDGEMENTS

Many people have been on this journey with me and contributed in many ways and I sincerely would like to thank all of you for guidance and support that I have obtained. Especially I would like to thank:

Giedre Grigelioniene, my main supervisor, for giving me the opportunity to do research – a journey that have been inspiring and educative. For enthusiastically pinpointing the important radiological clues to a skeletal dysplasia (even though I many times think they're all the same). Thanks for helping me to develop my scientific writing and all the support throughout this journey.

Magnus Nordenskjöld, my former main supervisor, now co-supervisor, for the possibility to combine clinical laboratory work with research. For your support, smart advices and always having an answer to my questions.

Ann Nordgren, my co-supervisor, for always asking relevant questions, being enthusiastic and supportive. For all the fun discussions and exchanges of ideas.

Daniel Nilsson, my co-supervisor for your support and expertise in bioinformatics and its shortcomings have been of great value for me.

Anna Lindstrand, my co-supervisor, for being inspiring and giving me encouragement and exchange of your experiences in life.

I also want to acknowledge present and former fellows of the research group **Rare Diseases**. Thanks for the nice atmosphere and sharing of knowledge and many interesting discussions about science and life in general: **Nina Jäntti**, for all your help and for all our little chats. **Sofia Frisk**, it's you and me ☺, thank you for being my training companion. **Raquel Vaz**, for being a good friend and bringing a fresh spirit to the lab. **Maria Pettersson**, you're the aCGH-guru, thank you for great collaboration during these years and even more to come. **Fulya Taylan**, thank you for our small talks and for assistance with analysis. **Benedicte Bang**, **Jesper Eisfeldt**, **Wolfgang Hofmeister** and **Bianca Tesi**, for good company.

My former room-mates: **Alice Costantini**, for all good advice, friendship and fun company at the course in Italy and **Anders Kämpe**, for your amazing ability to explain complicated things and making them understandable, including the weather.

A big hug to **Kicki Lagerstedt-Robinson** and **Helena Malmgren** always helpful and having answers to my questions. DNA-sektionen: **Eva Flöjs, Inger Malmberg, Monika Carlsson, Lena Forsberg, Lisa Stolpe, Isabel Neira, Cecilia Österman, Mikaela Davody-Bodell** and **Hilda Lundén Miguel**; for your positive attitude and making it so much fun working with you.

Irene White, Margareta Lagerberg and **Kristina Adolfsson** for good times and inspiration, who always encouraged me to go on every adventure, even this one.

All the people at the **Clinical Genetics** at Karolinska University Hospital. Just to mention a few: **Emma T, Josephine, Ellika, Christina N, Karin K, Anna T, Riffat, Fatemah, Malin H, Nina K, Weini, Camilla H, Johanna L, Malin K, Ahn-Nhi, Karin W, Lisen, Erik I, Britt-Marie, Agne** and **Paula** for your invaluable help and experience in clinical laboratory work. **Johanna R, Jenny vS, Johanna W, Ingegerd, Lena L, Hero, Erik B, Eva E, Cecilia R, Marita, Ingela, Kicki Å** and **Marlene** for nice discussions and your support.

Former and present collaborators at the research side: **Samina, Isabel T, Outi, Anna B, Anna-Lena, Clara, Miriam, Dominyka, Noor, Jessica, Anna W, Måns, Valtteri, Andrei, Phil, Nikos** and his lab team. The administration at MMK especially **Ann-Britt Wikström** and **Britt-Marie Witasp** for always quickly responding to my questions.

I would also like to express my deepest gratitude to the participating patients and their families, as well as to all collaborators involved in this research, both national and international. An extra thought to our skeletal dysplasia expert **Gen Nishimura**, thank you for your generous support in radiographic pattern recognition and tight collaboration in all my projects.

Finally, thanks to my friends and family, that have loaded my batteries and listened to the joy and concerns about my research.

Annika Ahlford a great mentor, as well the rest of the team from high school: **Magdalena Grebius, Charlotte Thörner, Maria Sjöstedt** and **Stina Bengtsson**, for good advices when I needed it the most. My friends from university ("Bondeskolan"): **Rebecka Pestoff, Cecilia Tingstedt, Lisa Sandberg, Nina Eriksson** and **Catharina Larsson**, from laughter to despair, with everything from sharing lecture notes to fredagsmys with marrängsviss and endless parties at Ultuna studentkår. Most of all thank you for still hanging in there, it has been more than twenty years since our first dates and looking forward to the rest!

Not to forget about **Johanna Nurminen**, my biggest supporter and dear friend from primary school, **Johanna Erikson**, giving us so much joy with **Minus** – *the cat* – and **Kattis, Marcus, Mats**, and **Isabell**, for friendship and all members of **pysselgruppen** and **innebandygänget** for having something else than research to think about!

My family: my mother **Kerstin** and **Bengt**, for taking care of the kids, all joy and late nights at Mörkö, my father **Claes** and **Anita**, for your support. My twin brother **Per** and **Sandra**, for great fun and wine, my supportive siblings **Anders**, **Göran** and **Eva**, and my extended family **Anne** and **Lasse**, for always having time to take care of the kids whenever we asked for help.

Edvin, **Albin** and **Ludvig** – my wonderful children, you can't imagine how much love and joy you give me in life. Thanks for all the fantastic adventures we have had and are going to have in life!

Anders, for all your love and support!

The work in this thesis has been supported by Karolinska Institutet Faculty Funds for Doctoral students (3-1233/2013), through the regional agreement on medical training and clinical research (ALF20150143) between Stockholm County Council and Karolinska Institutet, by grants from Stiftelsen Sällskapet Barnavård, Kronprinsessan Lovisas and Axel Tiellmans Minnesfond, Stiftelsen Promobilia and Stiftelsen Frimurare Barnhuset. I would also like to acknowledge support from Clinical Genomics, National Genomics Infrastructure (NGI), and UPPMAX (at Science for Life Laboratory) for assistance in massively parallel DNA sequencing and computational infrastructure, and Knut and Alice Wallenberg Foundation for the microscope facility.

7 REFERENCES

1 Bonafe, L. *et al.* Nosology and classification of genetic skeletal disorders: 2015 revision. *American journal of medical genetics. Part A* **167A**, 2869-2892, doi:10.1002/ajmg.a.37365 (2015).

2 Stevenson, D. A., Carey, J. C., Byrne, J. L., Srisukhumbowornchai, S. & Feldkamp, M. L. Analysis of skeletal dysplasias in the Utah population. *American journal of medical genetics. Part A* **158A**, 1046-1054, doi:10.1002/ajmg.a.35327 (2012).

3 Krakow, D. & Rimoin, D. L. The skeletal dysplasias. *Genet Med* **12**, 327-341, doi:10.1097/GIM.0b013e3181daae9b (2010).

4 Barbosa-Buck, C. O. *et al.* Clinical epidemiology of skeletal dysplasias in South America. *American journal of medical genetics. Part A* **158A**, 1038-1045, doi:10.1002/ajmg.a.35246 (2012).

5 Grigelioniene, G. *et al.* Extending the phenotype of lethal skeletal dysplasia type al Gazali. *American journal of medical genetics. Part A* **155A**, 1404-1408, doi:10.1002/ajmg.a.33990 (2011).

6 Chinen, Y. *et al.* Progressive hip joint subluxation in Saul-Wilson syndrome. *American journal of medical genetics. Part A* **167A**, 2834-2838, doi:10.1002/ajmg.a.37278 (2015).

7 Spranger, J. B., P.; Superti-Furga, A.; Unger, S.; Nishimura, G. *Bone Dysplasias - An atlas of Genetic Disorders of Skeletal Development*. 3rd edn, 828 (OUP USA, 2012).

8 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921, doi:10.1038/35057062 (2001).

9 Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304-1351, doi:10.1126/science.1058040 (2001).

10 International Human Genome Sequencing, C. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-945, doi:10.1038/nature03001 (2004).

11 Kidd, J. M. *et al.* Characterization of missing human genome sequences and copy-number polymorphic insertions. *Nat Methods* **7**, 365-371 (2010).

12 Elliott, T. A. & Gregory, T. R. What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. *Philos Trans R Soc Lond B Biol Sci* **370**, 20140331, doi:10.1098/rstb.2014.0331 (2015).

13 Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature genetics* **40**, 1413-1415, doi:10.1038/ng.259 (2008).

14 Salvatore, M., Magrelli, A. & Taruscio, D. The role of microRNAs in the biology of rare diseases. *Int J Mol Sci* **12**, 6733-6742, doi:10.3390/ijms12106733 (2011).

15 Huynh, N. P., Anderson, B. A., Guilak, F. & McAlinden, A. Emerging roles for long noncoding RNAs in skeletal biology and disease. *Connect Tissue Res* **58**, 116-141, doi:10.1080/03008207.2016.1194406 (2017).

16 Sera, S. R. & Zur Nieden, N. I. microRNA Regulation of Skeletal Development. *Curr Osteoporos Rep* **15**, 353-366, doi:10.1007/s11914-017-0379-7 (2017).

17 Amberger, J. S. & Hamosh, A. Searching Online Mendelian Inheritance in Man (OMIM): A Knowledgebase of Human Genes and Genetic Phenotypes. *Curr Protoc Bioinformatics* **58**, 1 2 1-1 2 12, doi:10.1002/cpbi.27 (2017).

18 Bonafe, L. *et al.* Nosology and classification of genetic skeletal disorders: 2015 revision. *American journal of medical genetics. Part A*, doi:10.1002/ajmg.a.37365 (2015).

19 Superti-Furga, A., Bonafe, L. & Rimoin, D. L. Molecular-pathogenetic classification of genetic disorders of the skeleton. *American journal of medical genetics* **106**, 282-293 (2001).

20 Warman, M. L. *et al.* Nosology and classification of genetic skeletal disorders: 2010 revision. *American journal of medical genetics. Part A* **155A**, 943-968, doi:10.1002/ajmg.a.33909 (2011).
21 Katsanis, N. *et al.* Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder. *Science* **293**, 2256-2259, doi:10.1126/science.1063525 (2001).
22 Lindstrand, A. *et al.* Copy-Number Variation Contributes to the Mutational Load of Bardet-Biedl Syndrome. *American journal of human genetics* **99**, 318-336, doi:10.1016/j.ajhg.2015.04.023 (2016).
23 Thiel, C. *et al.* NEK1 mutations cause short-rib polydactyly syndrome type Majewski. *American journal of human genetics* **88**, 106-114, doi:10.1016/j.ajhg.2010.12.004 (2011).
24 Schaffer, A. A. Digenic inheritance in medical genetics. *Journal of medical genetics* **50**, 641-652, doi:10.1136/jmedgenet-2013-101713 (2013).
25 Henderson, J. E. *et al.* Expression of FGFR3 with the G380R achondroplasia mutation inhibits proliferation and maturation of CFK2 chondrocytic cells. *J Bone Miner Res* **15**, 155-165, doi:10.1359/jbmr.2000.15.1.155 (2000).
26 Ridanpaa, M. *et al.* Mutations in the RNA component of RNase MRP cause a pleiotropic human disease, cartilage-hair hypoplasia. *Cell* **104**, 195-203 (2001).
27 Deng, H., Huang, X. & Yuan, L. Molecular genetics of the COL2A1-related disorders. *Mutat Res Rev Mutat Res* **768**, 1-13, doi:10.1016/j.mrrev.2016.02.003 (2016).
28 Reiter, J. F. & Leroux, M. R. Genes and molecular pathways underpinning ciliopathies. *Nat Rev Mol Cell Biol* **18**, 533-547, doi:10.1038/nrm.2017.60 (2017).
29 Schmidts, M. Clinical genetics and pathobiology of ciliary chondrodysplasias. *J Pediatr Genet* **3**, 46-94, doi:10.3233/PGE-14089 (2014).
30 Mitchison, H. M. & Valente, E. M. Motile and non-motile cilia in human pathology: from function to phenotypes. *J Pathol* **241**, 294-309, doi:10.1002/path.4843 (2017).
31 Song, B., Haycraft, C. J., Seo, H. S., Yoder, B. K. & Serra, R. Development of the post-natal growth plate requires intraflagellar transport proteins. *Dev Biol* **305**, 202-216, doi:10.1016/j.ydbio.2007.02.003 (2007).
32 Cortes, C. R. *et al.* Mutations in human C2CD3 cause skeletal dysplasia and provide new insights into phenotypic and cellular consequences of altered C2CD3 function. *Sci Rep* **6**, 24083, doi:10.1038/srep24083 (2016).
33 Bachmann-Gagescu, R. *et al.* Joubert syndrome: a model for untangling recessive disorders with extreme genetic heterogeneity. *Journal of medical genetics* **52**, 514-522, doi:10.1136/jmedgenet-2015-103087 (2015).
34 Roosing, S. *et al.* Mutations in CEP120 cause Joubert syndrome as well as complex ciliopathy phenotypes. *Journal of medical genetics* **53**, 608-615, doi:10.1136/jmedgenet-2016-103832 (2016).
35 Shaheen, R. *et al.* A founder CEP120 mutation in Jeune asphyxiating thoracic dystrophy expands the role of centriolar proteins in skeletal ciliopathies. *Human molecular genetics* **24**, 1410-1419, doi:10.1093/hmg/ddu555 (2015).
36 Melbouci, M. *et al.* Growth impairment in mucopolysaccharidoses. *Mol Genet Metab* **124**, 1-10, doi:10.1016/j.ymgme.2018.03.004 (2018).
37 Jaeken, J. & Peanne, R. What is new in CDG? *J Inherit Metab Dis* **40**, 569-586, doi:10.1007/s10545-017-0050-6 (2017).
38 Coman, D., Irving, M., Kannu, P., Jaeken, J. & Savarirayan, R. The skeletal manifestations of the congenital disorders of glycosylation. *Clinical genetics* **73**, 507-515, doi:10.1111/j.1399-0004.2008.01015.x (2008).
39 Ng, B. G. & Freeze, H. H. Perspectives on Glycosylation and Its Congenital Disorders. *Trends Genet* **34**, 466-476, doi:10.1016/j.tig.2018.03.002 (2018).

40 Weischenfeldt, J., Symmons, O., Spitz, F. & Korbel, J. O. Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat Rev Genet* **14**, 125-138, doi:10.1038/nrg3373 (2013).
41 Poggi, H. *et al.* A Deletion of More than 800 kb Is the Most Recurrent Mutation in Chilean Patients with SHOX Gene Defects. *Horm Res Paediatr* **84**, 254-257, doi:10.1159/000439109 (2015).
42 Benito-Sanz, S. *et al.* Identification of the first recurrent PAR1 deletion in Leri-Weill dyschondrosteosis and idiopathic short stature reveals the presence of a novel SHOX enhancer. *Journal of medical genetics* **49**, 442-450, doi:10.1136/jmedgenet-2011-100678 (2012).
43 Burkardt, D. D. *et al.* Distinctive phenotype in 9 patients with deletion of chromosome 1q24-q25. *American journal of medical genetics. Part A* **155A**, 1336-1351, doi:10.1002/ajmg.a.34049 (2011).
44 MacDonald, J. R., Ziman, R., Yuen, R. K., Feuk, L. & Scherer, S. W. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res* **42**, D986-992, doi:10.1093/nar/gkt958 (2014).
45 Lupianez, D. G., Spielmann, M. & Mundlos, S. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends Genet* **32**, 225-237, doi:10.1016/j.tig.2016.01.003 (2016).
46 Symmons, O. *et al.* The Shh Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances. *Dev Cell* **39**, 529-543, doi:10.1016/j.devcel.2016.10.015 (2016).
47 Klopocki, E. *et al.* A microduplication of the long range SHH limb regulator (ZRS) is associated with triphalangeal thumb-polysyndactyly syndrome. *Journal of medical genetics* **45**, 370-375, doi:10.1136/jmg.2007.055699 (2008).
48 van der Werf, I. M. *et al.* Novel microdeletions on chromosome 14q32.2 suggest a potential role for non-coding RNAs in Kagami-Ogata syndrome. *European journal of human genetics: EJHG* **24**, 1724-1729, doi:10.1038/ejhg.2016.82 (2016).
49 Turan, S. & Bastepe, M. GNAS Spectrum of Disorders. *Curr Osteoporos Rep* **13**, 146-158, doi:10.1007/s11914-015-0268-x (2015).
50 de Pontual, L. *et al.* Germline deletion of the miR-17 approximately 92 cluster causes skeletal and growth defects in humans. *Nature genetics* **43**, 1026-1030, doi:10.1038/ng.915 (2011).
51 McInerney-Leo, A. M. *et al.* Whole exome sequencing is an efficient, sensitive and specific method for determining the genetic cause of short-rib thoracic dystrophies. *Clinical genetics* **88**, 550-557, doi:10.1111/cge.12550 (2015).
52 Zhang, W. *et al.* Expanding the genetic architecture and phenotypic spectrum in the skeletal ciliopathies. *Human mutation* **39**, 152-166, doi:10.1002/humu.23362 (2018).
53 Zhang, H. *et al.* A pilot study of gene testing of genetic bone dysplasia using targeted next-generation sequencing. *J Hum Genet* **60**, 769-776, doi:10.1038/jhg.2015.112 (2015).
54 Caparros-Martin, J. A. *et al.* Molecular spectrum and differential diagnosis in patients referred with sporadic or autosomal recessive osteogenesis imperfecta. *Mol Genet Genomic Med* **5**, 28-39, doi:10.1002/mgg3.257 (2017).
55 Bardai, G., Moffatt, P., Glorieux, F. H. & Rauch, F. DNA sequence analysis in 598 individuals with a clinical diagnosis of osteogenesis imperfecta: diagnostic yield and mutation spectrum. *Osteoporos Int* **27**, 3607-3613, doi:10.1007/s00198-016-3709-1 (2016).
56 Saudi Mendeliome, G. Comprehensive gene panels provide advantages over clinical exome sequencing for Mendelian diseases. *Genome Biol* **16**, 134, doi:10.1186/s13059-015-0693-2 (2015).

57 Pang, A. W. *et al.* Towards a comprehensive structural variation map of an individual human genome. *Genome Biol* **11**, R52, doi:10.1186/gb-2010-11-5-r52 (2010).
58 Gomez, F., Hirbo, J. & Tishkoff, S. A. Genetic variation and adaptation in Africa: implications for human evolution and disease. *Cold Spring Harb Perspect Biol* **6**, a008524, doi:10.1101/cshperspect.a008524 (2014).
59 Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291, doi:10.1038/nature19057 (2016).
60 Ameur, A. *et al.* SweGen: a whole-genome data resource of genetic variability in a cross-section of the Swedish population. *European journal of human genetics: EJHG* **25**, 1253-1260, doi:10.1038/ejhg.2017.130 (2017).
61 Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).
62 Ng, P. C. & Henikoff, S. Predicting deleterious amino acid substitutions. *Genome Res* **11**, 863-874, doi:10.1101/gr.176601 (2001).
63 Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248-249, doi:10.1038/nmeth0410-248 (2010).
64 Davydov, E. V. *et al.* Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol* **6**, e1001025, doi:10.1371/journal.pcbi.1001025 (2010).
65 Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics* **46**, 310-315, doi:10.1038/ng.2892 (2014).
66 Besenbacher, S. *et al.* Novel variation and de novo mutation rates in population-wide de novo assembled Danish trios. *Nat Commun* **6**, 5969, doi:10.1038/ncomms6969 (2015).
67 Acuna-Hidalgo, R., Veltman, J. A. & Hoischen, A. New insights into the generation and role of de novo mutations in health and disease. *Genome Biol* **17**, 241, doi:10.1186/s13059-016-1110-1 (2016).
68 Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405-424, doi:10.1038/gim.2015.30 (2015).
69 Dorschner, M. O. *et al.* Actionable, pathogenic incidental findings in 1,000 participants' exomes. *American journal of human genetics* **93**, 631-640, doi:10.1016/j.ajhg.2013.08.006 (2013).
70 Wang, J. & Shen, Y. When a "disease-causing mutation" is not a pathogenic variant. *Clin Chem* **60**, 711-713, doi:10.1373/clinchem.2013.215947 (2014).
71 Smedley, D. *et al.* A Whole-Genome Analysis Framework for Effective Identification of Pathogenic Regulatory Variants in Mendelian Disease. *American journal of human genetics* **99**, 595-606, doi:10.1016/j.ajhg.2016.07.005 (2016).
72 Padgett, R. A. New connections between splicing and human disease. *Trends Genet* **28**, 147-154, doi:10.1016/j.tig.2012.01.001 (2012).
73 Ohno, K., Takeda, J. I. & Masuda, A. Rules and tools to predict the splicing effects of exonic and intronic mutations. *Wiley Interdiscip Rev RNA* **9**, doi:10.1002/wrna.1451 (2018).
74 Soemedi, R. *et al.* Pathogenic variants that alter protein code often disrupt splicing. *Nature genetics* **49**, 848-855, doi:10.1038/ng.3837 (2017).
75 Lee, Y. & Rio, D. C. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu Rev Biochem* **84**, 291-323, doi:10.1146/annurev-biochem-060614-034316 (2015).

76 Vaz-Drago, R., Custodio, N. & Carmo-Fonseca, M. Deep intronic mutations and human disease. *Hum Genet* **136**, 1093-1111, doi:10.1007/s00439-017-1809-4 (2017).
77 Popp, M. W. & Maquat, L. E. Organizing principles of mammalian nonsense-mediated mRNA decay. *Annu Rev Genet* **47**, 139-165, doi:10.1146/annurev-genet-111212-133424 (2013).
78 Posey, J. E. *et al.* Resolution of Disease Phenotypes Resulting from Multilocus Genomic Variation. *N Engl J Med* **376**, 21-31, doi:10.1056/NEJMoa1516767 (2017).
79 Balci, T. B. *et al.* Debunking Occam's razor: Diagnosing multiple genetic diseases in families by whole-exome sequencing. *Clinical genetics* **92**, 281-289, doi:10.1111/cge.12987 (2017).
80 Fareed, M. & Afzal, M. Genetics of consanguinity and inbreeding in health and disease. *Ann Hum Biol* **44**, 99-107, doi:10.1080/03014460.2016.1265148 (2017).
81 Alrukban, H. & Chitayat, D. Fetal chondrodysplasia punctata associated with maternal autoimmune diseases: a review. *Appl Clin Genet* **11**, 31-44, doi:10.2147/TACG.S150982 (2018).
82 Florencio-Silva, R., Sasso, G. R., Sasso-Cerri, E., Simoes, M. J. & Cerri, P. S. Biology of Bone Tissue: Structure, Function, and Factors That Influence Bone Cells. *Biomed Res Int* **2015**, 421746, doi:10.1155/2015/421746 (2015).
83 Long, F. & Ornitz, D. M. Development of the endochondral skeleton. *Cold Spring Harb Perspect Biol* **5**, a008334, doi:10.1101/cshperspect.a008334 (2013).
84 Akiyama, H. & Lefebvre, V. Unraveling the transcriptional regulatory machinery in chondrogenesis. *J Bone Miner Metab* **29**, 390-395, doi:10.1007/s00774-011-0273-9 (2011).
85 Karsenty, G., Kronenberg, H. M. & Settembre, C. Genetic control of bone formation. *Annu Rev Cell Dev Biol* **25**, 629-648, doi:10.1146/annurev.cellbio.042308.113308 (2009).
86 Ohba, S. Hedgehog Signaling in Endochondral Ossification. *J Dev Biol* **4**, doi:10.3390/jdb4020020 (2016).
87 Schipani, E., Kruse, K. & Juppner, H. A constitutively active mutant PTH-PTHrP receptor in Jansen-type metaphyseal chondrodysplasia. *Science* **268**, 98-100 (1995).
88 Loshkajian, A. *et al.* Familial Blomstrand chondrodysplasia with advanced skeletal maturation: further delineation. *American journal of medical genetics* **71**, 283-288 (1997).
89 Klopocki, E. *et al.* Deletion and point mutations of PTHLH cause brachydactyly type E. *American journal of human genetics* **86**, 434-439, doi:10.1016/j.ajhg.2010.01.023 (2010).
90 Gao, B. *et al.* Mutations in IHH, encoding Indian hedgehog, cause brachydactyly type A-1. *Nature genetics* **28**, 386-388, doi:10.1038/ng577 (2001).
91 Hellemans, J. *et al.* Homozygous mutations in IHH cause acrocapitofemoral dysplasia, an autosomal recessive disorder with cone-shaped epiphyses in hands and hips. *American journal of human genetics* **72**, 1040-1046 (2003).
92 Zheng, Q. *et al.* Dysregulation of chondrogenesis in human cleidocranial dysplasia. *American journal of human genetics* **77**, 305-312, doi:10.1086/432261 (2005).
93 Ikegawa, S. *et al.* Mutation of the type X collagen gene (COL10A1) causes spondylometaphyseal dysplasia. *American journal of human genetics* **63**, 1659-1662, doi:10.1086/302158 (1998).
94 Bellus, G. A. *et al.* Distinct missense mutations of the FGFR3 lys650 codon modulate receptor kinase activation and the severity of the skeletal dysplasia phenotype. *American journal of human genetics* **67**, 1411-1421, doi:10.1086/316892 (2000).

95 Baron, J. *et al.* Short and tall stature: a new paradigm emerges. *Nat Rev Endocrinol* **11**, 735-746, doi:10.1038/nrendo.2015.165 (2015).
96 Berendsen, A. D. & Olsen, B. R. Bone development. *Bone* **80**, 14-18, doi:10.1016/j.bone.2015.04.035 (2015).
97 Lapunzina, P. *et al.* Identification of a frameshift mutation in Osterix in a patient with recessive osteogenesis imperfecta. *American journal of human genetics* **87**, 110-114, doi:10.1016/j.ajhg.2010.05.016 (2010).
98 Mundlos, S. *et al.* Mutations involving the transcription factor CBFA1 cause cleidocranial dysplasia. *Cell* **89**, 773-779 (1997).
99 Chen, H. *et al.* Chondrocyte-specific regulatory activity of Runx2 is essential for survival and skeletal development. *Cells Tissues Organs* **194**, 161-165, doi:10.1159/000324743 (2011).
100 Jabs, E. W. *et al.* A mutation in the homeodomain of the human MSX2 gene in a family affected with autosomal dominant craniosynostosis. *Cell* **75**, 443-450 (1993).
101 Seto, M. L. *et al.* Isolated sagittal and coronal craniosynostosis associated with TWIST box mutations. *American journal of medical genetics. Part A* **143A**, 678-686, doi:10.1002/ajmg.a.31630 (2007).
102 Silva, B. C. & Bilezikian, J. P. Parathyroid hormone: anabolic and catabolic actions on the skeleton. *Curr Opin Pharmacol* **22**, 41-50, doi:10.1016/j.coph.2015.03.005 (2015).
103 Palagano, E., Menale, C., Sobacchi, C. & Villa, A. Genetics of Osteopetrosis. *Curr Osteoporos Rep* **16**, 13-25, doi:10.1007/s11914-018-0415-2 (2018).
104 Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat Biotechnol* **26**, 1135-1145, doi:10.1038/nbt1486 (2008).
105 Lazarus, S., Zankl, A. & Duncan, E. L. Next-generation sequencing: a frameshift in skeletal dysplasia gene discovery. *Osteoporos Int* **25**, 407-422, doi:10.1007/s00198-013-2443-1 (2014).
106 Sims, D., Sudbery, I., Ilott, N. E., Heger, A. & Ponting, C. P. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet* **15**, 121-132, doi:10.1038/nrg3642 (2014).
107 Gilissen, C., Hoischen, A., Brunner, H. G. & Veltman, J. A. Disease gene identification strategies for exome sequencing. *European journal of human genetics: EJHG* **20**, 490-497, doi:10.1038/ejhg.2011.258 (2012).
108 Belkadi, A. *et al.* Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A* **112**, 5473-5478, doi:10.1073/pnas.1418631112 (2015).
109 Stuppia, L., Antonucci, I., Palka, G. & Gatta, V. Use of the MLPA assay in the molecular diagnosis of gene copy number alterations in human genetic diseases. *Int J Mol Sci* **13**, 3245-3276, doi:10.3390/ijms13033245 (2012).
110 Bejjani, B. A. & Shaffer, L. G. Application of array-based comparative genomic hybridization to clinical diagnostics. *J Mol Diagn* **8**, 528-533, doi:10.2353/jmoldx.2006.060029 (2006).
111 Song, Z., Zhang, X., Jia, S., Yelick, P. C. & Zhao, C. Zebrafish as a Model for Human Ciliopathies. *J Genet Genomics* **43**, 107-120, doi:10.1016/j.jgg.2016.02.001 (2016).
112 Howe, K. *et al.* The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498-503, doi:10.1038/nature12111 (2013).
113 Kettleborough, R. N. *et al.* A systematic genome-wide analysis of zebrafish protein-coding gene function. *Nature* **496**, 494-497, doi:10.1038/nature11992 (2013).
114 Geisler, R., Borel, N., Ferg, M., Maier, J. V. & Strahle, U. Maintenance of Zebrafish Lines at the European Zebrafish Resource Center. *Zebrafish* **13 Suppl 1**, S19-23, doi:10.1089/zeb.2015.1205 (2016).

115 Walker, M. B. & Kimmel, C. B. A two-color acid-free cartilage and bone stain for zebrafish larvae. *Biotech Histochem* **82**, 23-28, doi:10.1080/10520290701333558 (2007).
116 Lacey, J. M., Bergen, H. R., Magera, M. J., Naylor, S. & O'Brien, J. F. Rapid determination of transferrin isoforms by immunoaffinity liquid chromatography and electrospray mass spectrometry. *Clin Chem* **47**, 513-518 (2001).
117 de la Morena-Barrio, M. E. *et al.* Hypoglycosylation is a common finding in antithrombin deficiency in the absence of a SERPINC1 gene defect. *J Thromb Haemost* **14**, 1549-1560, doi:10.1111/jth.13372 (2016).
118 Iqbal, Z. *et al.* Homozygous and heterozygous disruptions of ANK3: at the crossroads of neurodevelopmental and psychiatric disorders. *Human molecular genetics* **22**, 1960-1970, doi:10.1093/hmg/ddt043 (2013).
119 Hardies, K. *et al.* Recessive mutations in SLC13A5 result in a loss of citrate transport and cause neonatal epilepsy, developmental delay and teeth hypoplasia. *Brain* **138**, 3238-3250, doi:10.1093/brain/awv263 (2015).
120 Schossig, A. *et al.* SLC13A5 is the second gene associated with Kohlschutter-Tonz syndrome. *Journal of medical genetics* **54**, 54-62, doi:10.1136/jmedgenet-2016-103988 (2017).
121 Nishimura, G., Nakayama, M., Fuke, Y. & Suehara, N. A lethal osteochondrodysplasia with mesomelic brachymelia, round pelvis, and congenital hepatic fibrosis: two siblings born to consanguineous parents. *Pediatr Radiol* **28**, 43-47, doi:10.1007/s002470050289 (1998).
122 Gillessen-Kaesbach, G. *et al.* New autosomal recessive lethal disorder with polycystic kidneys type Potter I, characteristic face, microcephaly, brachymelia, and congenital heart defects. *American journal of medical genetics* **45**, 511-518, doi:10.1002/ajmg.1320450422 (1993).
123 Lepais, L. *et al.* ALG3-CDG: Report of two siblings with antenatal features carrying homozygous p.Gly96Arg mutation. *American journal of medical genetics. Part A* **167A**, 2748-2754, doi:10.1002/ajmg.a.37232 (2015).
124 Murali, C. *et al.* Diagnosis of ALG12-CDG by exome sequencing in a case of severe skeletal dysplasia. *Mol Genet Metab Rep* **1**, 213-219, doi:10.1016/j.ymgmr.2014.04.004 (2014).
125 Kranz, C. *et al.* Expanding spectrum of congenital disorder of glycosylation Ig (CDG-Ig): sibs with a unique skeletal dysplasia, hypogammaglobulinemia, cardiomyopathy, genital malformations, and early lethality. *American journal of medical genetics. Part A* **143A**, 1371-1378, doi:10.1002/ajmg.a.31791 (2007).
126 AlSubhi, S. *et al.* Further Delineation of the ALG9-CDG Phenotype. *JIMD Rep* **27**, 107-112, doi:10.1007/8904_2015_504 (2016).
127 Davis, K. *et al.* ALG9-CDG: New clinical case and review of the literature. *Mol Genet Metab Rep* **13**, 55-63, doi:10.1016/j.ymgmr.2017.08.004 (2017).
128 Al Teneiji, A. *et al.* Phenotypic and genotypic spectrum of congenital disorders of glycosylation type I and type II. *Mol Genet Metab* **120**, 235-242, doi:10.1016/j.ymgme.2016.12.014 (2017).
129 Stephen, J. *et al.* Mutations in KIAA0753 cause Joubert syndrome associated with growth hormone deficiency. *Hum Genet* **136**, 399-408, doi:10.1007/s00439-017-1765-z (2017).
130 Chevrier, V. *et al.* OFIP/KIAA0753 forms a complex with OFD1 and FOR20 at pericentriolar satellites and centrosomes and is mutated in one individual with oral-facial-digital syndrome. *Human molecular genetics* **25**, 497-513, doi:10.1093/hmg/ddv488 (2016).

131 Yuan, X. & Yang, S. Deletion of IFT80 Impairs Epiphyseal and Articular Cartilage Formation Due to Disruption of Chondrocyte Differentiation. *PloS one* **10**, e0130618, doi:10.1371/journal.pone.0130618 (2015).
132 Hoornaert, K. P. *et al.* Czech dysplasia metatarsal type: another type II collagen disorder. *European journal of human genetics: EJHG* **15**, 1269-1275, doi:10.1038/sj.ejhg.5201913 (2007).
133 Jurgens, J. *et al.* Novel COL2A1 variant (c.619G>A, p.Gly207Arg) manifesting as a phenotype similar to progressive pseudorheumatoid dysplasia and spondyloepiphyseal dysplasia, Stanescu type. *Human mutation* **36**, 1004-1008, doi:10.1002/humu.22839 (2015).
134 Hoornaert, K. P. *et al.* The phenotypic spectrum in patients with arginine to cysteine mutations in the COL2A1 gene. *Journal of medical genetics* **43**, 406-413, doi:10.1136/jmg.2005.035717 (2006).
135 Nishimura, G. *et al.* The phenotypic spectrum of COL2A1 mutations. *Human mutation* **26**, 36-43, doi:10.1002/humu.20179 (2005).
136 Barat-Houari, M. *et al.* Confirmation of autosomal recessive inheritance of COL2A1 mutations in spondyloepiphyseal dysplasia congenita: Lessons for genetic counseling. *American journal of medical genetics. Part A* **170A**, 263-265, doi:10.1002/ajmg.a.37374 (2016).
137 Nishimura, G. *et al.* Ischio-spinal dysostosis: a previously unrecognised combination of malformations. *Pediatr Radiol* **29**, 212-217, doi:10.1007/s002470050574 (1999).
138 Legare, J. M., Seaborg, K., Laffin, J. & Giampietro, P. F. Diaphanospondylodysostosis and ischiospinal dysostosis, evidence for one disorder with variable expression in a patient who has survived to age 9 years. *American journal of medical genetics. Part A* **173**, 2808-2813, doi:10.1002/ajmg.a.38395 (2017).
139 Kremer, L. S. *et al.* Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat Commun* **8**, 15824, doi:10.1038/ncomms15824 (2017).
140 Cummings, B. B. *et al.* Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med* **9**, doi:10.1126/scitranslmed.aal5209 (2017).
141 Chakravorty, S. & Hegde, M. Clinical Utility of Transcriptome Sequencing: Toward a Better Diagnosis for Mendelian Disorders. *Clin Chem* **64**, 882-884, doi:10.1373/clinchem.2017.276980 (2018).
142 Stranneheim, H. *et al.* Rapid pulsed whole genome sequencing for comprehensive acute diagnostics of inborn errors of metabolism. *BMC Genomics* **15**, 1090, doi:10.1186/1471-2164-15-1090 (2014).
143 Philippakis, A. A. *et al.* The Matchmaker Exchange: a platform for rare disease gene discovery. *Human mutation* **36**, 915-921, doi:10.1002/humu.22858 (2015).
144 Sobreira, N. L. M. *et al.* Matchmaker Exchange. *Curr Protoc Hum Genet* **95**, 9 31 31-39 31 15, doi:10.1002/cphg.50 (2017).
145 Smits, P. *et al.* Lethal skeletal dysplasia in mice and humans lacking the golgin GMAP-210. *N Engl J Med* **362**, 206-216, doi:10.1056/NEJMoa0900158 (2010).
146 Green, R. C. *et al.* ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med* **15**, 565-574, doi:10.1038/gim.2013.73 (2013).
147 Padhi, D., Jang, G., Stouch, B., Fang, L. & Posvar, E. Single-dose, placebo-controlled, randomized study of AMG 785, a sclerostin monoclonal antibody. *J Bone Miner Res* **26**, 19-26, doi:10.1002/jbmr.173 (2011).

148 Lorget, F. *et al.* Evaluation of the therapeutic potential of a CNP analog in a Fgfr3 mouse model recapitulating achondroplasia. *American journal of human genetics* **91**, 1108-1114, doi:10.1016/j.ajhg.2012.10.014 (2012).